Journal of Experimental Psychology:
Learning, Memory, and Cognition
1998, Vol. 24, No. 4, 803-821

0278-7393/98/$3.00

Word Naming and Memory Load: Still Searching for an Individual Differences Explanation

Penny M. Pexman and Stephen J. Lupker
University of Western Ontario

S. E. Bernstein and T. H. Carr (1996) and P. M. Pexman and S. J. Lupker (1995) suggested that classifiable individual differences in word-naming performance can account for the varied findings on the naming and memory load task (NMLT; K. R. Paap & R. W. Noel, 1991). Bernstein and Carr's technique of testing their explanation by using performance on part of the NMLT to classify participants is problematic, however. To remedy this, in the present study participants were classified on the basis of performance on a priori tasks: Participants completed a naming task, a naming task with low memory load, and the NMLT. Performance on the NMLT was not predicted by performance on either a priori task, thus providing no support for either Bernstein and Carr's or Pexman and Lupker's individual differences accounts.

In recent years, a considerable amount of research effort has been directed at understanding what appears to be a simple process—the process of pronouncing words. Although this process itself has proven to be much more complex than initially thought, the research has produced a number of relatively consistent empirical findings. For example, word-naming latencies tend to be faster for high-frequency words than for low-frequency words (e.g., Andrews, 1982; Brown, Lupker, & Colombo, 1994; Forster & Chambers, 1973). In addition, naming latencies for irregular words, words for which the spelling-to-sound relationship is not standard (e.g., *sword*), are typically longer than latencies for regular words (e.g., *sweet;* Baron & Strawson, 1976). Further, this "regularity effect" has been found to interact with word frequency (Andrews, 1982; Brown et al., 1994; Paap & Noel, 1991; Seidenberg, Waters, Barnes, & Tanenhaus, 1984; Taraban & McClelland, 1987); that is, whereas low-frequency irregular words are named more slowly than low-frequency regular words, no regularity effect has typically been found for high-frequency words (although see Content, 1991; Jared, 1995; and Lupker, Brown, & Colombo, 1997).

This interaction of frequency and regularity was initially taken as strong support for dual-route theories of naming (Coltheart, 1978; see also the more recent dual-route cascaded model of Coltheart, Curtis, Atkins, & Haller, 1993). According to the dual-route hypothesis, words are represented as entries in a mental lexicon. These entries can be accessed by a direct look-up process. That is, a word's orthographic information can be used by the reader to directly access the word's lexical entry, thus enabling essentially holistic retrieval of the word's phonological code. This route is referred to as the "lexical route." The second route, referred to as the "assembly route," is a rule-based process for deriving a pronunciation by using subword units. This route operates in parallel with the lexical route. In dual-route models in which the two routes are independent (e.g., Paap & Noel, 1991), the pronunciation produced by the reader essentially comes from whichever route is the first to generate a usable phonological code.

Dual-route theorists account for the interaction of frequency and regularity by assuming that (a) the lexical route is frequency sensitive in that processing speed is a direct function of frequency and (b) naming times are prolonged if the two routes produce different phonological codes at approximately the same time. For example, Paap and Noel (1991) described processing by the two routes as a horse race where the code produced by the faster route is declared the winner, and it is this code that is produced. If, however, the two routes produce different pronunciations at approximately the same time, although the code from the lexical route tends ultimately to be used, a time-consuming decision process is undertaken to resolve the competition between routes.

Editor's Note. The appearance of this article was delayed, with the authors' permission, so that it could appear in the same issue as the following two articles by Bernstein, DeShon, and Carr and by Paap and Herdman.—JHN

Penny M. Pexman and Stephen J. Lupker, Department of Psychology, University of Western Ontario, London, Ontario, Canada.

This research was supported by a graduate fellowship to Penny M. Pexman and by Grant A6333, both from the Natural Sciences and Engineering Research Council of Canada. Much of the research described in this article was also presented at the 36th Annual Meeting of the Psychonomic Society, Los Angeles, California, in November 1995. We thank Stuart Bernstein, Tom Carr, Bob Gardner, Ken Paap, Max Coltheart, and Debra Jared for their many helpful discussions about this research.

Correspondence concerning this article should be addressed to Penny M. Pexman or Stephen J. Lupker, Department of Psychology, University of Western Ontario, London, Ontario N6A 5C2 Canada. Electronic mail may be sent to Penny M. Pexman at pexman@sscl.uwo.ca or to Stephen J. Lupker at lupker@julian.uwo.ca.

Competition can arise, of course, only when naming irregular words. That is, although irregular words (e.g., *pint*) are processed accurately by the lexical route, processing by the assembly route produces an erroneous regularized pronunciation. Further, competition will arise for irregular words only if those words are sufficiently low in frequency that they are processed slowly enough by the lexical route to allow a competition situation to arise. In contrast, low-frequency regular words (e.g., *tale*) are pronounced relatively quickly because the pronunciation generated by either route is correct. The regularity effect for low-frequency words is, then, a result of the assembly route harming irregular words while at the same time helping regular words. On the other hand, for high-frequency words there is no regularity effect because the lexical route produces the correct pronunciation for all of these words quite rapidly, thus avoiding any influence from the assembly route.

Recently, a number of challenges to the dual-route model have emerged based on the premise that only a single route is necessary for accurate pronunciation of all letter strings (Brown & Besner, 1987; Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg & McClelland, 1989; Van Orden, Pennington, & Stone, 1990). In Plaut et al.'s (1996) model, for example, the process of word naming is modelled by the creation of activation patterns in sets of orthographic and phonological units. Between sets of orthographic and phonological units, there is a set of hidden units. Connections between all units are weighted, and the connection weights are changed as the model goes through a training (i.e., learning) process in order to best represent the spelling-to-sound correspondences for the words to which the model is being exposed.

In this model, the effects of frequency and regularity can be explained in terms of the connection weights that result from training. The important point is that during the training process, the connection weights are adjusted slightly to better reflect the spelling-to-sound correspondences of the word the system has just encountered. The result is that the weights come to better reflect the spelling-to-sound correspondences of words seen more frequently. Thus, high-frequency words would be expected to be named more quickly than low-frequency words because, as a result of the training process, the system has come to represent their spelling-to-sound correspondences better than the correspondences for low-frequency words. Similarly, regular words are named more quickly than irregular words because the model encounters the spelling-to-sound patterns of regular words more often (i.e., in many similar words) than it encounters those of irregular words. Simulations of the model demonstrate that it also can account for the Frequency × Regularity interaction (Plaut et al., 1996; Seidenberg & McClelland, 1989).

The ability of single-route models, like Plaut et al.'s (1996) model, to account for the Frequency × Regularity interaction has caused researchers to look elsewhere for evidence of the existence of two independent routes to pronunciation. For instance, Monsell, Patterson, Graham, Hughes, and Milroy (1992) attempted to demonstrate that participants could strategically control the use of the assembly route in a naming task, thus supporting the existence of such a route. However, Lupker et al. (1997) obtained results that suggest that Monsell et al.'s results are better explained by other types of strategic adjustments—adjustments that do not necessitate a dual-route account of naming performance.

Another contribution to the debate between dual- and single-route models was a study by Paap and Noel (1991). Paap and Noel produced new evidence of separate lexical and assembly routes for word naming by using a paradigm that combined word naming with a memory task.

The premise of Paap and Noel's (1991) experiment was that the lexical and assembly processes for naming have different attentional requirements. Both routes are assumed to require attentional resources to derive pronunciations, but the assembly route is held to be more affected when attentional resources are limited. Processing on the assembly route is believed to require more resources for operation because it is an active process involving assembly of phonological codes. Conversely, the lexical route is held to be more automatic because it relies on a direct mapping from orthography to lexical units.

The hypothesis of Paap and Noel (1991) is that when attentional resources are limited because of a high memory load, the attention-demanding assembly route should slow down more than the lexical route. The result would be that naming would be driven mainly by the lexical route. As such, naming of low-frequency irregular words would be much less likely to be affected by competition from the assembly route, and those words might actually be named *more quickly* under a high memory load than a low memory load. On the other hand, with no memory load the assembly route often provides the correct phonological code for low-frequency regular words. Thus, these words would be quite negatively affected by a high memory load if it meant that the assembly route was generally not available to support the lexical route. In essence, the prediction was that the regularity effect for low-frequency words should be either greatly reduced or eliminated under high memory load because the naming latencies of low-frequency regular and irregular words become very similar.

A second effect of a high memory load is that it should produce a large frequency effect for regular words. Under normal circumstances, the frequency effect is quite small for regular words. The presumed reason is, as mentioned above, that the assembly route is assumed to assist to a large extent in the pronunciation of low-frequency regular words. With the imposition of a high memory load, this assistance would be substantially reduced, and the size of the frequency effect would increase so that it would be virtually as large as that for irregular words.

These predictions were clearly supported by the results of Paap and Noel's (1991) study. They found a significant three-way interaction of frequency, regularity, and memory load such that the size of the regularity effect for low-frequency words decreased dramatically under the high memory load, whereas the size of the frequency effect for regular words increased dramatically. These results were taken as providing reasonably strong new support for the dual-route position.

Before proceeding, we would like to make an additional point about Paap and Noel's (1991) experiment. Although Paap and Noel's results have been taken as reasonably strong new support for the dual-route model, failure to find those results would not have discredited the model. That is, the predictions being examined were based on a number of additional assumptions generated by Paap and Noel about the attentional demands made by the two routes and how a load manipulation would affect them. These assumptions are not part of the basic dual-route framework, and if they had been incorrect, the results could have turned out quite differently without having any implications for the model itself.

Because of the importance of Paap and Noel's (1991) results in the dual-route versus single-route debate, since the publication of Paap and Noel's article, a number of researchers have attempted to replicate their findings (e.g., Herdman, Beckett, & Stolpmann, 1993; D. Jared, personal communication, March 1993; E. Strain, personal communication, April 1993). In one of the two previous attempts most central to the present article, Pexman and Lupker (1995) reported three experiments involving a direct replication of Paap and Noel's Experiment 1 procedure (Pexman & Lupker, 1995, Experiments 1A–1C). Unfortunately, in none of these experiments was there any hint of a replication of Paap and Noel's pattern (i.e., a reduction in the regularity effect for low-frequency words and an increase in the frequency effect for regular words in the high-memory-load condition). In Experiment 2, Pexman and Lupker increased the high memory load to seven digits (Paap & Noel, 1991, had used five) but again failed to produce Paap and Noel's pattern. Because another group of researchers (Herdman et al., 1993) had reported successfully replicating Paap and Noel's results by using a different set of stimuli, in a final experiment, Pexman and Lupker used this alternative stimulus set. Once again, however, they failed to replicate Paap and Noel's results.

In attempting to explain their inability to replicate Paap and Noel's (1991) results, Pexman and Lupker (1995) suggested an explanation based on individual differences in word-naming ability. This account was based on the subsymbolic, single-route model of Van Orden et al. (1990). Van Orden et al. suggested that there is a "clean-up" component of phonological-code generation during which an initially "noisy" encoding is cleaned up to match an "attractor" (the correct phonological code). This situation can be modelled by assuming that the initial noisy encoding and the attractor represent two points in a phonological space. The time it takes the clean-up process to reach completion is assumed to be a function of the distance of the initial encoding from the attractor. To this explanation Pexman and Lupker added the assumption that clean-up time is a function of the degree to which there are other attractors that are legitimate competitors. For low-frequency irregular words, for instance, legitimate competition might arise from the regularized version of the word's phonology, a situation that would prolong the clean-up process.

According to this explanation, in Paap and Noel's (1991) task, the effect of increased memory load is to hinder the ability of all competitors to compete. This could have two possible results. On the one hand, if there are no legitimate competitors, the result would be simply a slowing down of the clean-up process. This would lead to slightly longer naming latencies overall but not necessarily to any change in the size of the regularity effect. On the other hand, if there are many somewhat legitimate competitors, the result could be a freeing of the eventual winner from the effects of competition. Possibly, this could even lead to a speedup of the clean-up process for low-frequency irregular words depending on how strong the competitors are.

The extension of this explanation to the difference between Paap and Noel's (1991) results and Pexman and Lupker's (1995) results was that Paap and Noel's participants might have experienced more competition among possible encodings for low-frequency irregular words than did Pexman and Lupker's participants, for whom the process of naming low-frequency irregular words may have been qualitatively similar to the process of naming all other types of words. This idea gains some support from the fact that Paap and Noel's participants did show a much larger low-frequency regularity effect in the one-digit (low-memory-load) condition than Pexman and Lupker's did, suggesting that in general, they had much more difficulty with low-frequency irregular words.

In the other replication attempt most central to the present article, Bernstein and Carr (1996) not only used the digit-load task but also used memory loads of nouns, random shapes, and pseudowords. For digit and noun loads, they produced a weak version of Paap and Noel's (1991) pattern when using Paap and Noel's stimuli; however, their random-shape and pseudoword memory loads did not show the same pattern, despite the fact that these loads were found to be more attention demanding. Further, none of the loads produced Paap and Noel's pattern when a different set of words was named.

Bernstein and Carr (1996) also pursued an individual differences explanation for their findings and suggested that skilled readers do not all use a dual-route system for word naming. That is, although the naming process for some of their participants could be described by a dual-route architecture, others named words in a way characteristic of a single route.

To test this explanation, Bernstein and Carr (1996) selected participants whose naming process appeared to be characteristic of a dual-route architecture on the basis of their naming performance in the low-memory-load condition of Paap and Noel's (1991) task. That is, they specified selection criteria designed to identify those participants whose naming latencies fit the typical Frequency × Regularity interaction and then applied those criteria to performance in the low-load condition of Paap and Noel's task. When only those participants who met these criteria were considered, their naming performance in the high-memory-load condition was very similar to the performance of Paap and Noel's participants. Thus, it seemed, Paap and Noel's effects were replicable, but only for a certain type of reader.

Unfortunately, there is a problem with Bernstein and Carr's (1996) individual differences analysis due to the fact that they selected participants by using a part of the very task for which they were trying to predict performance. In particular, they selected a set of participants who already showed half of the desired three-way interaction (i.e., participants who showed quite large regularity effects in the low-load condition). Thus, the strength of support for their individual differences explanation is somewhat compromised. That is, although it is possible that their explanation is correct, it is also possible that these participants showed a reduced regularity effect in the high-load condition because of a regression-to-the-mean type of artifact. Clearly, Bernstein and Carr's argument for their individual differences explanation would have been substantially stronger if they had used their criteria to select participants on a task that was separate from the one for which they were trying to predict performance.

Bernstein and Carr (1996) were not unaware of this potential problem and, in fact, attempted to address it by carrying out a general linear model (GLM) analysis on the data from all their participants. Effectively, this analysis creates a continuous independent variable out of the participants' low- and high-frequency regularity effects in the one-digit-load condition and includes that variable as a factor in the analysis of naming latencies in both the one- and five-digit-load conditions. Bernstein and Carr argued that because this analysis did not involve any selection of participants it would not be influenced by a regression-to-the-mean type of artifact. In this analysis, the four-way interaction involving this new variable and the variables of frequency, regularity, and memory load was significant, which Bernstein and Carr took as evidence that their results were not an artifact of their participant-selection technique but were due to real individual differences.

It is somewhat unclear to us, however, why the GLM analysis would guard against the regression-to-the-mean problem that we have noted. The one-digit-load performance variable (the new independent variable in the GLM analysis) is basically also the measure that drives Bernstein and Carr's (1996) selection criteria. Thus, in the GLM analysis, this new variable still involves performance on part of the task for which they were trying to predict performance. As such, its use essentially still builds in half of the desired interaction.

More importantly, it is quite easy to rule out this regression-to-the-mean explanation and to support Bernstein and Carr's (1996) idea of classifiable individual differences in the way people name words. One simply needs to demonstrate that it is possible to apply their criteria to naming performance in a task *external* to Paap and Noel's (1991) task and still successfully select participants who will show Paap and Noel's pattern when tested in Paap and Noel's task.

The main goal for the present article is to provide an evaluation of these individual differences explanations. Because Bernstein and Carr's (1996) criteria were much better specified than Pexman and Lupker's (1995), the analysis focuses mainly on Bernstein and Carr's account. To accomplish this, participants were presented with three tasks in the following order—a simple naming task (no-load naming), a naming task with a low (one-digit) memory load for all trials (low-load naming), and Paap and Noel's (1991) task, which is a naming task involving a low (one-digit) memory load for half of the trials and and a high (five-digit) memory load for the remainder of the trials (criterion task). The first two tasks allowed us to select participants who fulfilled the criteria specified by Bernstein and Carr, so that we could examine those participants' performance in the third task—Paap and Noel's task.

Participants were selected on the basis of their performance in no-load naming and, separately, on the basis of performance in low-load naming because it was unclear which task would function better as a predictor of criterion task performance. For example, there is the possibility that simply having a memory load might affect naming performance in some unspecified way. Thus, participants selected on the basis of performance in no-load naming may not be the ones most likely to show Paap and Noel's pattern in the criterion task. On the other hand, because the one-digit load is at least somewhat attention demanding, in low-load naming there may be a decrease in the size of the regularity effect for the true dual-route participants. As such, it may be more difficult to identify them on the basis of this task than on the basis of a task having no extra attentional demands.

With respect to testing Pexman and Lupker's (1995) individual differences explanation (which was based on the extent to which different people experience competition from other viable encodings during naming), as noted, the selection criteria for testing this explanation are much more vague than Bernstein and Carr's (1996). Thus, the approach taken here was more exploratory. We considered several different criteria all relating to the issue of the amount of competition an individual might experience when processing irregular words.

Method

Participants

The participants were 50 (29 female) undergraduate students at the University of Western Ontario who were given partial course credit in an Introductory Psychology course in exchange for their participation. The average age of participants was 20.60 years. All participants considered English to be their native language and had normal or corrected-to-normal vision.

Apparatus

An IBM-clone Trillium Computer Resources PC (Model No. 316S-80MS) was used to control the presentation of stimuli. Stimuli were presented on a TTX Multiscan Monitor (Model No. 3435P), for which letters and digits were approximately 0.60 cm high and at eye level for participants. The distance between the participants and the monitor screen was approximately 50 cm. Naming response times were recorded by a Shure, Inc. (Model No. 5755) microphone connected to a Gerbrands (Model No. 800) electronic voice key relay. Participants were instructed to sit approximately 20 cm from the microphone. The *shift* keys on a standard keyboard were used to make responses in the memory task in low-load naming and in the criterion task. These keys were labeled *YES* and *NO* such that a participant's dominant hand was assigned to the *YES* key.

Stimuli

No-load naming. No-load naming was the first task presented to all participants. The stimuli used in this task consisted of 10 words for use in the practice trials, in addition to 15 low-frequency irregular words (mean frequency = 14, median frequency = 12, mean neighborhood size [MNBS] = 7), 15 low-frequency regular words (mean frequency = 10, median frequency = 8, MNBS = 11), 15 high-frequency irregular words (mean frequency = 665, median frequency = 213, MNBS = 8), and 15 high-frequency regular words (mean frequency = 590, median frequency = 390, MNBS = 9) for use in the experimental trials. For all words, these reported frequencies were based on Kučera and Francis's (1967) norms.

As much as possible for stimuli in no-load naming and low-load naming, each irregular word was matched with a regular word for frequency and first phoneme. The words for no-load naming and low-load naming were selected from the regular and irregular word sets used by Herdman et al. (1993), Seidenberg et al. (1984), and Taraban and McClelland (1987).[1]

Low-load naming. Low-load naming was the second task presented to all participants. The stimuli used in this task consisted of 10 practice words in addition to 15 low-frequency irregular words (mean frequency = 11, median frequency = 6, MNBS = 7), 15 low-frequency regular words (mean frequency = 7, median frequency = 5, MNBS = 10), 15 high-frequency irregular words (mean frequency = 2,000, median frequency = 108, MNBS = 7), and 15 high-frequency regular words (mean frequency = 1,501, median frequency = 447, MNBS = 9) for use in the experimental trials.

Criterion task. This task was the final task for all participants. The stimuli used in this task were the same words used in Paap and Noel's (1991) Experiment 1. These stimuli consisted of 20 practice words in addition to 20 low-frequency irregular words (mean frequency = 3.5, MNBS = 8), 20 low-frequency regular words (mean frequency = 3.5, MNBS = 10), 20 high-frequency irregular words (mean frequency = 393, MNBS = 9), and 20 high-frequency regular words (mean frequency = 214, MNBS = 11) for use in the experimental trials.[2] The regularity of these words was evaluated by Paap and Noel, and their article should be consulted for details. In creating the word list, Paap and Noel matched each irregular word with a regular word for approximate frequency, initial phoneme, and length. The stimuli used in all tasks are presented in the Appendix.

Procedure

For the criterion task, as in Paap and Noel's (1991) study, half of the words in each of the Frequency × Regularity cells were presented to participants in the one-digit-memory-load condition, and half were presented in the five-digit-memory-load condition. Therefore, two groups of participants were needed to counterbalance assignment of words to load conditions. Participants were assigned to these groups by order of participation in the study such that the odd-numbered participants were assigned to one group, and the even-numbered participants were assigned to the other group. There were 25 participants in each group.

Participants were tested individually, and each completed all three tasks. With short breaks between tasks, the testing session lasted approximately 55 min. The testing room was normally lit. In all three tasks, each trial began with a 50-ms 400-Hz beep signal. Following the beep, a fixation point appeared on the screen for 1 s.

For no-load naming, participants were instructed to name each word as quickly and accurately as they could as soon as it appeared on the monitor. For low-load naming and the criterion task, participants were instructed that they were to perform both a naming task and a memory task. They were asked to remember that the memory task was primary and the naming task secondary. Participants were asked to make their naming responses as quickly and accurately as possible and to make their memory responses as accurately as possible.

Each trial in low-load naming and the criterion task included both the memory task and the naming task. The sequence of each trial in low-load naming and the criterion task was identical, except that in the criterion task there were two memory-load conditions (one vs. five digits to remember) and in low-load naming there was only one condition (one digit to remember). That trial sequence was as follows: When the initial fixation point disappeared, the digit or digits to remember would appear on the monitor for 2 s. The digits for each trial were randomly selected. The monitor was then blank for a variable interval of either 1 or 2 s. After this interval, the word to be named appeared on the screen. When the voice key was activated by the response, the word disappeared. Six seconds after the start of the trial, one digit appeared on the screen. Participants were instructed to make a "yes" or "no" response to this digit, according to whether the digit was the digit (or one of the digits) that had appeared at the start of the trial. The "yes" or "no" response was made by pressing one of the two *shift* keys on the keyboard.

For each of the three tasks, the experimenter sat behind the participant and recorded any naming errors that were made.

Results and Discussion

In all cases, data were analyzed with subjects (F_s) and items (F_i) as separate random variables. Thus, values for both types of analyses are reported. The criterion for statistical significance was $p < .05$. Trials on which naming

[1] There is, of course, some disagreement as to how to determine whether a word is irregular. In the present experiment, we simply accepted these authors' classification schemes. We are grateful to Max Coltheart for providing us, post hoc, with a classification of these words according to his classification scheme (Coltheart et al., 1993). As one might expect, according to this classification scheme, a few of our stimuli would be classified differently. In particular, for no-load naming stimuli, *BROW* and *PLOW*, which we used as irregular words, were classified as regular words, whereas *GROUP*, which we used as a regular word, was classified as an irregular word according to Coltheart et al.'s (1993) classification scheme. For low-load naming, *CEASE*, *DOLL*, *SPOOK*, and *SHALL*, which we used as irregular words, were classified as regular words according to Coltheart et al.'s classification scheme.

To determine whether these stimuli had any effect on the results reported in this article, we reran every analysis with these words excluded. The results of the analyses performed without these items did not differ significantly (or even marginally) from the results of the analyses reported in the article.

[2] The mean frequencies of the stimuli used in this study, especially for high-frequency words, varied across the three tasks. In part, this was because we had very little choice for high-frequency irregular words and, thus, could not control mean frequency for those words as well as we would have liked. The fact that the median frequencies for each of our types of words in each task were reasonably similar led us to conclude that the words were matched sufficiently well for present purposes.

latencies were faster than 250 ms or slower than 1,400 ms, or trials on which the voice key was not triggered, were not analyzed. These trials made up fewer than 1% of all trials. In addition, trials on which participants made naming errors were excluded from the analyses.

Naming Latencies: No-Load Naming

The mean naming latencies for each of the four conditions in the no-load naming task are presented in Table 1. For these data, the interaction of frequency and regularity was significant, $F_s(1, 49) = 16.66$, $MSE = 467.53$; $F_i(1, 56) = 3.73$, $MSE = 1{,}424.86$, $p < .10$, as were the main effects of frequency, $F_s(1, 49) = 84.39$, $MSE = 1{,}298.43$; $F_i(1, 56) = 24.43$, $MSE = 1{,}424.86$, and regularity, $F_s(1, 49) = 75.37$, $MSE = 715.15$; $F_i(1, 56) = 12.02$, $MSE = 1{,}424.86$. Simple main effects tests showed that the regularity effect for low-frequency words was significant, $t(49) = 6.00$. The regularity effect for high-frequency words, although smaller, was also significant when applying a one-tailed criterion, $t(49) = 1.81$. Although this latter effect is not typical, it is not unprecedented (Content, 1991; Jared, 1995). More importantly, the stimuli chosen for this task produced the typical Frequency × Regularity interaction in which the regularity effect was significantly larger for low-frequency words.

Naming Errors: No-Load Naming

For naming errors in this task, 63% could be classified as regularizations where the participant, for instance, would pronounce *SOOT* as *SUIT.* The remainder of the naming errors were generally mispronunciations where the participant, for instance, would pronounce *BUNT* as *BURNT.* Mean error percentages for each condition are listed in Table 1. As with the latency data, analyses of the error data showed a significant interaction of frequency and regularity, $F_s(1, 49) = 22.29$, $MSE = 19.06$; $F_i(1, 56) = 2.82$, $MSE = 51.59$, $p < .15$, as well as main effects of frequency, $F_s(1, 49) = 54.42$, $MSE = 21.36$; $F_i(1, 56) = 8.49$, $MSE = 51.59$, and regularity, $F_s(1, 49) = 55.78$, $MSE = 24.42$; $F_i(1, 56) = 7.47$, $MSE = 51.59$.

Table 1
Mean Naming Latencies (in Milliseconds) and Mean Naming Errors (in Percentages) in No-Load Naming and Low-Load Naming

Word frequency and regularity	No-load naming		Low-load naming	
	M	% error	*M*	% error
Low frequency				
Irregular	590	10.0	647	10.2
Regular	556	1.9	604	1.1
Regularity effect	34	8.1	43	9.1
High frequency				
Irregular	540	2.9	590	0.9
Regular	530	0.0	570	0.0
Regularity effect	10	2.9	20	0.9

Table 2
Mean Naming Latencies (in Milliseconds) and Mean Naming Errors (in Percentages) in the Criterion Task

Word frequency and regularity	Memory load			
	1 digit		5 digits	
	M	% error	*M*	% error
Low frequency				
Irregular	642	12.7	641	11.6
Regular	592	0.6	600	0.2
Regularity effect	50	12.1	41	11.4
High frequency				
Irregular	592	1.4	595	1.8
Regular	577	0.4	589	0.4
Regularity effect	15	1.0	6	1.4

Naming Latencies: Low-Load Naming

The mean naming latencies from low-load naming, in which participants named words with a one-digit memory load, are also listed in Table 1. For these data, the interaction of frequency and regularity was significant, $F_s(1, 49) = 6.64$, $MSE = 363.82$; $F_i(1, 56) = 4.45$, $MSE = 1{,}449.61$, as were the main effects of frequency, $F_s(1, 49) = 80.10$, $MSE = 955.78$; $F_i(1, 56) = 20.45$, $MSE = 1{,}449.61$, and regularity, $F_s(1, 49) = 47.79$, $MSE = 484.37$; $F_i(1, 56) = 7.41$, $MSE = 1{,}449.61$. Simple main effects tests showed that the regularity effects for both low-frequency, $t(49) = 7.75$, and high-frequency, $t(49) = 3.82$, words were significant.

Naming Errors: Low-Load Naming

For naming errors in this task, 69% were regularization errors. Mean naming error percentages for each condition are presented in Table 1. As with the latency data, the analyses of these data showed an interaction of frequency and regularity, $F_s(1, 49) = 34.98$, $MSE = 23.91$; $F_i(1, 56) = 7.92$, $MSE = 32.30$, as well as main effects of frequency, $F_s(1, 49) = 60.36$, $MSE = 22.20$; $F_i(1, 56) = 12.10$, $MSE = 32.30$, and regularity, $F_s(1, 49) = 51.79$, $MSE = 24.46$; $F_i(1, 56) = 12.54$, $MSE = 32.30$.

Naming Latencies: Criterion Task

The mean naming latencies from the criterion task, in which participants named words under one-digit and five-digit memory loads, are presented in Table 2. As illustrated in Table 2, the regularity effects observed for low-frequency words under low (50 ms) and high (41 ms) memory loads created a pattern very different from those reported by Paap and Noel (1991; 80 ms and −11 ms for low and high memory loads, respectively).

As in no-load naming and low-load naming, the interaction of frequency and regularity, $F_s(1, 49) = 34.77$, $MSE = 865.35$; $F_i(1, 76) = 5.28$, $MSE = 4{,}069.41$, and the main effects of frequency, $F_s(1, 49) = 64.76$, $MSE = 1{,}446.35$; $F_i(1, 76) = 12.07$, $MSE = 4{,}069.41$, and regularity, $F_s(1, 49) = 48.62$, $MSE = 1{,}586.20$; $F_i(1, 76) = 11.25$, $MSE = 4{,}069.41$, were significant. No other effects approached significance. In particular, there was no hint of a three-way interaction of frequency, regularity, and memory load or a main effect of memory load ($F_s < 1$ and $F_i < 1$ in both cases).

Naming Errors: Criterion Task

For naming errors in this task, 82% were regularization errors. For each condition, mean naming error percentages are presented in Table 2. For these data, analyses also showed that the three-way interaction of frequency, regularity, and memory load was nonsignificant ($F_s < 1$, $F_i < 1$). The main effect of memory load was also nonsignificant ($F_s < 1$, $F_i < 1$). As with the latency data, the only significant effects were the interaction of frequency and regularity, $F_s(1, 49) = 138.29$, $MSE = 20.08$; $F_i(1, 76) = 4.28$, $MSE = 257.05$, and the main effects of frequency, $F_s(1, 49) = 147.28$, $MSE = 18.93$; $F_i(1, 76) = 4.28$, $MSE = 257.05$, and regularity, $F_s(1, 49) = 197.48$, $MSE = 21.41$; $F_i(1, 76) = 6.46$, $MSE = 257.05$.

Therefore, once again, using Paap and Noel's (1991) stimuli and experimental paradigm, we failed to replicate Paap and Noel's pattern. In fact, the present results are very similar to the other failures to replicate reported by Pexman and Lupker (1995).

As noted, Bernstein and Carr (1996) also had only minimal success replicating Paap and Noel's (1991) effects in their overall analyses. However, they were able to obtain Paap and Noel's pattern with participants who met their selection criteria. As such, we first set out to determine whether application of Bernstein and Carr's selection criteria to performance in the one-digit-memory-load condition of our criterion task would produce a subset of participants showing Paap and Noel's pattern. Those criteria were, first, "(a) high-frequency words were named faster than low-frequency words, (b) for low-frequency words, regular words were named faster than exception words, and (c) this difference was larger than the analogous difference observed for high-frequency words" (Bernstein & Carr, 1996, p. 100). Second, participants must be above the median in "a median split based on the difference in naming latencies between low-frequency regular words and low-frequency exception words under low memory load" (Bernstein & Carr, 1996, p. 100).

There were 21 participants (42%), in the present experiment, who met both of these criteria. Recall that to counterbalance words with memory-load conditions, two experimental groups had been created for this study. Of the selected participants, 8 had been in the first experimental group, and 13 had been in the second group. When Bernstein and Carr (1996) applied these criteria to their data, they selected 40% of the participants in their Experiment 2 and also 40% in their Experiment 3. Therefore, the proportion of participants that we selected was virtually identical to the proportion selected by Bernstein and Carr.

For our selected participants, naming latencies on the criterion task were analyzed. These latencies are presented in Table 3. The important three-way interaction of frequency, regularity and memory load was now significant, $F_s(1, 20) = 11.33$, $MSE = 1,145.72$; $F_i(1, 76) = 4.35$, $MSE = 1,431.91$. Other significant effects were the Frequency × Regularity interaction, $F_s(1, 20) = 102.18$, $MSE = 379.17$; $F_i(1, 76) = 8.93$, $MSE = 6,565.95$, and the main effects of frequency, $F_s(1, 20) = 61.99$, $MSE = 1,190.94$; $F_i(1, 76) = 15.30$, $MSE = 6,565.95$, and regularity, $F_s(1, 20) = 43.07$, $MSE = 828.45$; $F_i(1, 76) = 7.22$, $MSE = 6,565.95$. No other effects, including memory load, approached significance.

Table 3
Mean Naming Latencies (in Milliseconds) and Mean Naming Errors (in Percentages) for Participants Selected From the Criterion Task, Using Bernstein and Carr's (1996) Criteria

Word frequency and regularity	Memory load: 1 digit		5 digits	
	M	% error	*M*	% error
Low frequency				
Irregular	717	13.3	680	12.6
Regular	632	1.0	645	0.5
Regularity effect	85	12.3	35	12.1
High frequency				
Irregular	615	1.4	638	2.8
Regular	626	1.0	629	1.0
Regularity effect	−11	0.4	9	1.8

The naming error data for these 21 participants are also presented in Table 3. These error data showed the same effects as were found for naming latencies, except that the interaction of frequency, regularity, and memory load was not significant, $F_s(1, 20) = 1.48$, $MSE = 29.81$; $F_i(1, 76) = 1.29$, $MSE = 37.89$.

As Bernstein and Carr (1996) reported, it appears that Paap and Noel's (1991) complete pattern can be found for the subset of participants selected when applying Bernstein and Carr's criteria to performance in the one-digit-load condition. Thus, these data are a clear replication of Bernstein and Carr's results. The point to keep in mind, however, is that these participants were selected on the basis of their performance on a part of the task for which we were trying to predict performance. In particular, the criteria guaranteed that we would select participants who had the largest regularity effect for low-frequency words in the low-memory-load condition. Thus, there is always the possibility that for these participants, the decrease in the size of the regularity effect in the high-load condition may have been due to nothing more than a regression-to-the-mean type of artifact.

Bernstein and Carr's (1996) Individual Differences Explanation

Recall that Bernstein and Carr (1996) specified criteria for selecting participants based on the idea that there are certain participants who name words in ways that are characteristic of a dual-route architecture, and as a result, these people should show Paap and Noel's (1991) complete pattern. The argument here is that a valid test of their individual differences account would involve applying their criteria to select true dual-route participants on the basis of performance in an a priori task or tasks and then demonstrating that those participants showed the appropriate performance in Paap and Noel's task. As such, we had participants complete two tasks (no-load naming and low-load naming) prior to the criterion task. Either of these two tasks would allow us to apply Bernstein and Carr's criteria to select true dual-route participants.

Applying the criteria to naming latencies in no-load naming allowed us to select 23 of 50 participants (46%). Of the 23 participants selected on the basis of no-load naming latencies, 16 were from the first experimental group, and 7 were from the second group. Applying the criteria to low-load naming latencies, 23 participants were also selected. Here, 12 participants were from the first experimental group, and 11 participants were from the second group. It is interesting to note that only 15 of these participants had also been selected on the basis of no-load naming performance. If the selection criteria were essentially selecting participants randomly, the expected number of people who would meet the criteria in both tasks would be 10 or 11 (i.e., $.46 \times .46 \times 50 = 10.58$). Although 15 is slightly larger than this number, it does not differ significantly from it ($z = 1.53$), thus indicating that there was not a strong tendency for these criteria to select the same participants across the two, presumably, comparable tasks.

One conclusion that could be drawn from these results is that naming words under a one-digit memory load is not qualitatively the same task as naming words under no memory load. Indeed, low-load naming latencies were, on average, 50 ms slower than latencies in no-load naming. An analysis of naming latencies from no-load naming and low-load naming together, however, did not turn up any qualitative differences between tasks. That is, although the task effect was significant, $F_s(1, 49) = 52.79$, $MSE = 4{,}538.80$; $F_i(1, 112) = 47.17$, $MSE = 1{,}437.24$, the task effect did not interact with either regularity or frequency effects (or their interaction). Thus, it is unclear why the overlap between the two selected samples was not significantly greater than would be expected by chance. Because it is also unclear which task should be regarded as superior, participants selected on the basis of each task were used in separate analyses, and finally, an analysis was performed for the 15 participants who met the criteria in both tasks.

The mean naming latencies in the criterion task for the 23 participants selected on the basis of no-load-naming performance are presented in Table 4. For these participants, the interaction of frequency and regularity, $F_s(1, 22) = 17.45$, $MSE = 965.20$; $F_i(1, 76) = 5.76$, $MSE = 4{,}368.11$, and the main effects of frequency, $F_s(1, 22) = 53.16$, $MSE = 1{,}372.04$; $F_i(1, 76) = 14.41$, $MSE = 4{,}368.11$, and regularity, $F_s(1, 22) = 44.85$, $MSE = 1{,}234.64$; $F_i(1, 76) = 12.20$, $MSE = 4{,}368.11$, were significant. The three-way interaction of frequency, regularity, and memory load, however, did not approach significance, $F_s(1, 22) = 1.83$, $MSE = 2{,}267.74$; $F_i < 1$.

Although the three-way interaction was not significant for these individuals, their regularity effects did change somewhat from low to high memory load. In particular, their regularity effect for low-frequency words increased from 48 ms to 60 ms, and their regularity effect for high-frequency words decreased from 29 ms to 2 ms. A possible implication of our failure to observe a significant interaction is that there was very little power in this analysis to detect the interaction. In fact, with an alpha level of .05, the power to detect an interaction the size of Paap and Noel's (1991) was only .66. The point here, however, is that the change we did observe for the low-frequency regularity effect was actually in the *opposite direction* from that observed by Paap and Noel. Thus, it is clear that the participants selected on the basis of the application of Bernstein and Carr's (1996) criteria to no-load naming performance did not show Paap and Noel's pattern in the criterion task.

For these 23 participants selected on the basis of no-load naming performance, the naming error percentages were also analyzed and are presented in Table 4. The results of analyses for naming errors were the same as those for naming latencies. In particular, the interaction of frequency, regularity, and memory load was not significant, $F_s(1, 22) = 1.94$, $MSE = 56.81$; $F_i < 1$.

Table 4
Mean Naming Latencies (in Milliseconds) and Mean Naming Errors (in Percentages) in the Criterion Task for Participants Selected From No-Load Naming and Low-Load Naming, Using Bernstein and Carr's (1996) Criteria

Word frequency and regularity	No load				Low load			
	1 digit		5 digits		1 digit		5 digits	
	M	% error	*M*	% error	*M*	% error	*M*	% error
Low frequency								
Irregular	645	15.3	660	16.5	653	15.6	648	13.5
Regular	597	1.7	600	2.2	602	1.5	598	2.0
Regularity effect	48	13.6	60	14.3	51	14.1	50	11.5
High frequency								
Irregular	602	3.0	585	2.2	596	1.7	583	2.1
Regular	573	1.6	583	0.9	573	1.2	579	0.9
Regularity effect	29	1.4	2	1.3	23	0.5	4	1.2

The mean naming latencies in the criterion task for those participants selected on the basis of low-load naming performance are also presented in Table 4. Once again, the interaction of frequency and regularity, $F_s(1, 22) = 27.56$, $MSE = 568.98$; $F_i(1, 76) = 5.95$, $MSE = 6{,}119.13$, and the main effects of frequency, $F_s(1, 22) = 68.74$, $MSE = 1{,}232.03$; $F_i(1, 76) = 12.16$, $MSE = 6{,}119.13$, and regularity, $F_s(1, 22) = 43.01$, $MSE = 1{,}416.73$; $F_i(1, 76) = 7.12$, $MSE = 6{,}119.13$, were significant. However, the three-way interaction of frequency, regularity, and memory load did not approach significance ($F_s < 1$, $F_i < 1$). Note in particular that the size of the low-frequency regularity effect for these individuals was virtually unchanged from low to high memory load. Thus, the participants selected on the basis of the application of Bernstein and Carr's (1996) criteria to low-load naming performance did not show Paap and Noel's (1991) pattern in the criterion task either.

For these 23 participants selected on the basis of low-load naming performance, the naming error percentages were also analyzed and are presented in Table 4. The results were the same as those for naming latencies. In particular, the interaction of frequency, regularity, and memory load was not significant, $F_s(1, 22) = 1.85$, $MSE = 49.55$; $F_i < 1$.

As noted, there were 15 participants whose performance met Bernstein and Carr's (1996) selection criteria in no-load naming and also in low-load naming. Their mean criterion task latencies are presented in Table 5. For these participants, the interaction of frequency and regularity was significant, $F_s(1, 14) = 20.08$, $MSE = 494.20$; $F_i(1, 76) = 5.09$, $MSE = 4{,}516.33$, as were the main effects of frequency, $F_s(1, 14) = 41.35$, $MSE = 1{,}396.01$; $F_i(1, 76) = 15.11$, $MSE = 4{,}516.33$, and regularity, $F_s(1, 14) = 37.53$, $MSE = 840.11$; $F_i(1, 76) = 9.70$, $MSE = 4{,}516.33$. However, the three-way interaction of frequency, regularity, and memory load again did not approach significance ($F_s < 1$, $F_i < 1$). Note, in particular, that the size of the low-frequency regularity effect *increased,* rather than decreased, under high memory load. Thus, when participants were selected on the basis of both no-load naming and low-load naming performance, according to Bernstein and Carr's (1996) criteria, their criterion task naming latencies did not show the pattern observed by Paap and Noel (1991).

Table 5
Mean Naming Latencies (in Milliseconds) and Mean Naming Errors (in Percentages) in the Criterion Task for Participants Selected From Both No-Load Naming and Low-Load Naming, Using Bernstein and Carr's (1996) Criteria

	Memory load			
	1 digit		5 digits	
Word frequency and regularity	*M*	% error	*M*	% error
Low frequency				
Irregular	645	16.0	651	14.0
Regular	600	0.0	594	2.6
Regularity effect	45	16.0	57	11.4
High frequency				
Irregular	601	2.0	571	1.3
Regular	569	2.0	574	0.7
Regularity effect	32	0.0	−3	0.6

For these 15 participants, the naming error percentages were also analyzed and are presented in Table 5. The results of analyses for naming errors were the same as those for naming latencies. In particular, the interaction of frequency, regularity, and memory load was not significant ($F_s < 1$, $F_i < 1$).

The quite different patterns contained in Table 3 (where Bernstein & Carr's, 1996, criteria were used to select participants on the basis of performance in the one-digit load of Paap & Noel's, 1991, task) and Tables 4 and 5 (where the criteria were applied to the a priori tasks) also suggests that quite different participants had been selected in the two situations. Such was clearly the case: For the 23 participants selected on the basis of no-load naming performance, 9 of these participants had been among the 21 participants selected on the basis of their performance in the one-digit-memory-load condition of the criterion task. If we were selecting randomly, we would expect an overlap here of 9 or 10 participants ($.46 \times .42 \times 50 = 9.66$). For the 23 participants selected on the basis of low-load naming performance, 11 were also among the 21 selected on the basis of performance in the one-digit-memory-load condition of the criterion task. The expected random overlap in this case was also 9 or 10 participants. Finally, for the 15 participants selected on the basis of performance in both no-load naming and low-load naming, 6 of these participants were among the 21 selected on the basis of performance in the one-digit-memory-load condition of the criterion task. This value also matches the expected random overlap of these two groups ($.30 \times .42 \times 50 = 6.30$). Therefore, even though the selection overlap between no-load naming and low-load naming might be slightly (although not significantly) higher than expected by chance, the complete picture suggests that Bernstein and Carr's (1996) criteria selected different, and essentially random, groups of participants on the different tasks. This suggests that the different criteria are not identifying reliable individual differences, with the implication being that Bernstein and Carr likely obtained Paap and Noel's (1991) pattern of data for their special subgroup because of a regression-to-the-mean type of artifact.

The present results, therefore, provide little support for Bernstein and Carr's (1996) argument that there are classifiable differences in the way individuals name words (with some individuals consistently performing as if they had a dual-route architecture) and that their selection criteria allow researchers to identify those individuals. The present results do not, of course, rule out the possibility that some people do have dual-route architectures and some people do not. In fact, they do not even rule it out for the participants within Bernstein and Carr's experiments. No data that we could collect from participants in our experiments could tell us specifically about the mental architecture of the participants in Bernstein and Carr's experiments (or any other experiments for that matter). Nonetheless, on the basis of our results, it does seem more likely that the ability of Bernstein and Carr's criteria to predict criterion task performance, when their criteria were used to select participants on the basis of performance in the one-digit-load condition in that task, was artifactual in both their experiments and in the present experiment. That is, because the participants who were selected were those who showed the largest low-frequency regularity effects in the one-digit-load condition, it appears that the decrease in their regularity effect in the five-digit-load condition represented essentially a regression-to-the-mean type of artifact.

Another way of thinking about our account of why Bernstein and Carr's (1996) selection criteria tend to work when selecting participants on the basis of performance in the criterion task can be derived from starting with the assumption that all the irregular words used in the criterion task (or any task, for that matter) are not of the same difficulty for all participants. For some participants, the set of irregular low-frequency words selected to be used in their one-digit-load condition likely would be, on average, more difficult than the set selected to be used in their five-digit-load condition. The same applies to the low-frequency regular words. That is, they would not be of the same difficulty for all participants, and hence, for some participants, the set of words used in their one-digit-load condition would be easier than the set of words used in their five-digit-load condition. Bernstein and Carr's criteria, when applied to performance in the one-digit-load condition in Paap and Noel's (1991) task, tend to select those people who have more difficult irregular words and less difficult regular words in their one-digit-load condition than in their five-digit-load condition. The result for the selected people will be a strong tendency for there to be a smaller regularity effect in the five-digit-load condition than in the one-digit-load condition.

Pexman and Lupker's (1995) Individual Differences Explanation

Pexman and Lupker (1995) have argued that a possible individual differences explanation for mixed results with Paap and Noel's (1991) task might be based on the extent to which participants experience competition from viable competitors during their phonological-code-generation process. In particular, the participants who experience strong competition should be those who show Paap and Noel's pattern of a decreased low-frequency regularity effect under high memory load. Because the criteria for selecting those types of participants are not nearly as clear as Bernstein and Carr's (1996) criteria, four different criteria were investigated. In all cases, the criteria were applied separately to no-load naming and low-load naming performance, and the question in each case was whether the criteria would predict criterion task performance.

In our first attempt to test this idea, we decided to select those participants who showed the largest regularity effect for high-frequency words. The rationale was that high-frequency words are seen so often that their spelling-to-sound mappings must be very well learned. Thus, any participants showing a regularity effect for these words must be experiencing competition. We conducted a tertile split on the size of participants' high-frequency regularity effects in no-load naming and, separately, in low-load naming to select the one third of participants with the largest high-frequency regularity effects in both tasks. We then looked at how those two groups of participants performed in the criterion task. The regularity effects for these participants in the criterion task are presented in Table 6.

In the criterion task performance for this subset of

Table 6
Regularity Effects (Response Times, RTs, in Milliseconds and Error Percentages) in the Criterion Task for Pexman and Lupker's (1995) Individual Differences Analyses

	Effect size							
	Low frequency (LF)				High frequency (HF)			
	1 digit		5 digits		1 digit		5 digits	
Individual difference measure	RT	% error	RT	% error	RT	% error	RT	% error
HF regularity effect (n = 17)								
No-load naming	64	12.6	61	9.1	18	1.2	20	1.2
Low-load naming	63	13.2	61	10.1	24	0.8	20	0.9
LF regularity effect (n = 17)								
No-load naming	58	12.3	60	13.8	35	1.3	4	1.3
Low-load naming	48	14.5	51	12.6	13	1.4	10	1.6
Both LF and HF regularity effects								
No-load naming (n = 13)	73	16.1	64	15.6	42	1.7	22	1.9
Low-load naming (n = 11)	63	14.4	41	15.1	−2	0.6	29	0.9
LF irregular latency (n = 17)								
No-load naming	63	13.4	57	15.1	18	1.4	20	1.3
Low-load naming	70	14.7	46	15.4	9	0.7	22	1.7
LF irregular words with regularized pronunciation (n = 17)								
No-load naming	66	11.6	56	12.7	22	0.0	18	0.2
Low-load naming	60	12.4	48	13.1	12	1.1	23	1.1

participants, there was no hint of the three-way interaction of frequency, regularity, and memory load ($F_s < 1$ and $F_i < 1$ for naming latencies and naming error percentages for both the no-load naming and low-load naming groups). (The same was true when these analyses were carried out using median rather than tertile splits based on participants' high-frequency regularity effects.) Thus, this criterion did not allow us to select participants who showed Paap and Noel's (1991) pattern in the criterion task.

In a second attempt to test our individual differences explanation, we selected participants with relatively large low-frequency regularity effects. The rationale was simply that low-frequency irregular words would be more likely to be affected by competition than high-frequency irregular words, and thus, competition effects may be most evident for these words. (Note that this criterion selects a set of participants very similar to that selected by Bernstein & Carr's, 1996, criteria.) Again, we selected the top one third of participants on the basis of their latencies in no-load naming and separately in low-load naming. The regularity effects in the criterion task for these participants are also presented in Table 6. For these two groups of participants, the three-way interaction of frequency, regularity, and memory load in the criterion task was nonsignificant ($F_s < 1$ and $F_i < 1$ in all cases for both latencies and errors). (The same was true when these analyses were carried out using median rather than tertile splits based on participants' low-frequency regularity effects.) Thus, this criterion also did not allow us to select participants who showed Paap and Noel's (1991) pattern in the criterion task.[3]

The third criterion examined was the speed of participants' naming latencies for low-frequency irregular words. For all of our low-frequency irregular words, there is an incorrect regularized pronunciation that might be generated. If participants were particularly slow in responding to low-frequency irregular words, it is possibly because they were taking time to resolve competition between the words' irregular and regular pronunciations. We selected the slowest one third of participants for the low-frequency irregular words in no-load naming and separately in low-load naming. We then examined the criterion task performance for these two groups. The regularity effects in the criterion task for these participants are also presented in Table 6. Again, the three-way interaction of frequency, regularity, and memory load was nonsignificant ($F_s < 1.5$ and $F_i < 1.5$ for both no-load naming and low-load naming groups for both latencies and errors).

The final criterion that we used to select participants was the naming latency for low-frequency irregular words for which the regularized pronunciation of the word was also a real word (e.g., *BOWL* whose regularized pronunciation sounds like the word *BOWEL*). The rationale was that these types of words would be most likely to generate serious competition for readers because the competitor has a representation in phonological space by virtue of the fact that it also is a word. Thus, slow response times to these types of items may be an indicator of which participants were actually experiencing competition from alternate encodings. Among the no-load naming stimuli, there were four such words, and among the low-load naming stimuli there were six such words. We selected the one third of participants who were slowest to name these four words in no-load naming and then separately for the six words in low-load naming. The regularity effects in the criterion task for these participants are presented in Table 6. When we looked at criterion task performance for these two groups of participants, their three-way interaction of frequency, regularity, and memory load was nonsignificant ($F_s < 1$ and $F_i < 1$ for both no-load naming and low-load naming groups for both latencies and errors). Thus, none of the four criteria that we used to examine Pexman and Lupker's (1995) individual differences explanation for varied naming performance under high memory loads could predict which participants would show Paap and Noel's (1991) pattern in the criterion task.

Because neither Bernstein and Carr's (1996) selection criteria nor any of ours identified participants a priori who would show Paap and Noel's (1991) pattern, our final analysis involved looking directly at criterion task latencies and identifying two groups of participants: one group that best showed the desired pattern and one group that clearly did not. We then compared the performance of these groups on the a priori tasks to see if there were any detectable differences. To this end, for all participants' criterion task latencies, we calculated a value based on the following formula: (LFI1 − LFR1) − (LFI5 − LFR5), where LF is low frequency, I is irregular, R is regular, 1 is one-digit load, and 5 is five-digit load. This equation calculates the size of participants' low-frequency regularity effect in the one-digit-memory-load condition minus the size of their low-frequency regularity effect in the five-digit-memory-load condition. We ranked participants on the basis of the number we calculated when we applied this formula to their criterion task latencies, and classified the one third of participants ($n = 17$) with the highest values as showing the results most like Paap and Noel's (Group 1) and the one third of participants ($n = 17$) with the lowest values as showing the results least like Paap and Noel's (Group 2). For these groups, we analyzed criterion task latencies, including groups as a between-subjects variable in the analysis, to ensure that our classification had been effective. The mean latencies for each group on the criterion task are presented in Table 7.

The four-way interaction of frequency, regularity, memory load, and group was significant, $F_s(1, 32) = 17.94$, $MSE = 1{,}413.52$, $F_i(1, 76) = 5.65$, $MSE = 2{,}418.39$. Also, the interaction of frequency, regularity, and group was signifi-

[3] We also investigated the magnitude of participants' low-frequency *and* high-frequency regularity effects as a predictor of criterion task performance. We selected participants who fell above the median in terms of both their low-frequency and their high-frequency regularity effects. For no-load naming, 13 participants met this criterion, and for low-load naming, 11 participants met this criterion. For both these groups, their criterion task performance also showed a nonsignificant three-way interaction of frequency, regularity, and memory load ($F_s < 1$ and $F_i < 1$ in both cases for both latencies and errors). The regularity effects for these participants are also presented in Table 6.

Table 7
Mean Naming Latencies (in Milliseconds) and Mean Naming Errors (in Percentages) in the Criterion Task for Group 1 (Participants Who Showed Effects Most Like Paap & Noel's, 1991, Effects) and Group 2 (Participants Who Showed Effects Least Like Paap & Noel's Effects)

Word frequency and regularity	Group 1 (n = 17)				Group 2 (n = 17)			
	1 digit		5 digits		1 digit		5 digits	
	M	% error	*M*	% error	*M*	% error	*M*	% error
Low frequency								
Irregular	696	12.9	652	13.4	602	14.6	645	13.8
Regular	608	0.6	633	2.3	594	2.4	578	1.2
Regularity effect	88	12.3	19	11.1	8	12.2	67	12.6
High frequency								
Irregular	613	2.3	625	3.5	596	2.9	569	2.9
Regular	607	1.2	620	1.8	556	2.3	559	2.4
Regularity effect	6	1.1	5	1.7	40	0.6	10	0.5

cant, $F_s(1, 32) = 6.02$, $MSE = 936.21$; $F_i(1, 76) = 6.95$, $MSE = 1{,}501.50$, as was the interaction of frequency and regularity, $F_s(1, 32) = 16.64$, $MSE = 936.21$; $F_i(1, 76) = 3.75$, $MSE = 10{,}159.67$, $p < .10$, and the main effects of frequency, $F_s(1, 32) = 44.56$, $MSE = 1{,}625.49$; $F_i(1, 76) = 13.37$, $MSE = 10{,}159.67$, and regularity, $F_s(1, 32) = 30.13$, $MSE = 2{,}043.63$; $F_i(1, 76) = 10.03$, $MSE = 10{,}159.67$. The main effect of group was highly significant by items but not by subjects, $F_s(1, 32) = 2.49$, $MSE = 54{,}270.22$; $F_i(1, 76) = 97.32$, $MSE = 1{,}501.50$. No other effects were significant. The nature of these group effects indicates that our division of participants into appropriate groups was successful.

Mean naming errors (in percentages) are also presented in Table 7. The analysis of naming errors showed results slightly different from those in the analysis of latencies. None of the effects involving the group variable reached significance. In particular, the four-way interaction of group, frequency, regularity, and memory load was not significant, $F_s(1, 32) = 1.19$, $MSE = 41.56$; $F_i < 1$.

We then analyzed no-load naming and low-load naming data for these two groups of participants to see if there were any differences in their performance. The analysis included frequency, regularity, and task as within-subjects variables and group as a between-subjects variable. There was a significant interaction of frequency and regularity, $F_s(1, 32) = 16.13$, $MSE = 550.67$; $F_i(1, 112) = 2.75$, $MSE = 3{,}669.33$, $p < .15$, and main effects of frequency, $F_s(1, 32) = 121.39$, $MSE = 1{,}414.11$; $F_i(1, 112) = 50.57$, $MSE = 3{,}669.33$, regularity, $F_s(1, 32) = 78.47$, $MSE = 856.03$; $F_i(1, 112) = 16.48$, $MSE = 3{,}669.33$, and task, $F_s(1, 32) = 28.09$, $MSE = 5{,}718.73$; $F_i(1, 112) = 34.47$, $MSE = 3{,}669.33$. The nature of the task effect for these individuals was the same as that reported earlier when task differences were analyzed for all 50 participants. That is, naming latencies were longer in low-load naming (here by 49 ms), because a one-digit memory load slows naming latencies.

With respect to the question of group differences, there was a main effect of group that was significant by items but not by subjects, $F_s(1, 32) = 1.17$, $MSE = 55{,}453.00$; $F_i(1, 112) = 69.63$, $MSE = 732.72$. This main effect of group was due to Group 1 participants having an overall latency 30 ms slower than Group 2 participants. More importantly, however, group did not interact with any of the other effects (all F_ss < 1.45, all F_is < 2.10).

For analyses of naming errors, the effects were the same as those observed for naming latencies. That is, group did not interact significantly with any of the other variables in the analysis (all F_ss < 1, all F_is < 1).

The fact that in this analysis, there were no significant interactions with the group variable indicates that the groups did not differ on any of the variables measured, except for overall response time. Thus, with the exception of overall response time, there was no aspect of performance on either no-load naming or low-load naming that was different for the participants who showed Paap and Noel's (1991) pattern and those who did not.

The significant effect of group on overall response time does raise the possibility that overall response time might be a predictor of which participants would show Paap and Noel's (1991) pattern of results in the criterion task. To test this possibility, we selected the one third of participants who had the longest overall response times in no-load naming and separately selected the one third of participants who had the longest overall response times in low-load naming. When we analyzed criterion task performance for these participants, neither the subset selected from no-load naming nor the subset selected from low-load naming showed Paap and Noel's pattern ($F_s < 1$ and $F_i < 1$ for the three-way interaction for both subsets for both latencies and errors). Therefore, although Group 1 participants (those who showed Paap and Noel's pattern in the criterion task) had slower overall response times than Group 2 participants, overall response time on a priori tasks did not predict who would show Paap and Noel's pattern either.

General Discussion

The results of Paap and Noel's (1991) study appear to provide nice support for dual-route models of naming (although, as Pexman & Lupker, 1995, have noted, it is possible to provide an interpretation of these results within a single-route framework). Unfortunately, those results have proven to be difficult to replicate. The purpose of the present article was to investigate the individual differences accounts proposed by Bernstein and Carr (1996) and Pexman and Lupker (1995) to explain these replication failures. In the end, there was very little support for either of these accounts.

As we have demonstrated, using Bernstein and Carr's (1996) criteria to select participants on the basis of performance in the one-digit-load condition of Paap and Noel's (1991) task does produce a group of participants who show Paap and Noel's complete pattern. This procedure is problematic, however. The problem is that the participants selected are the ones who already show half of the three-way interaction and may then show the rest of the interaction because of a regression-to-the-mean type of artifact. To address this problem, we used the same criteria to select participants on the basis of performance in a priori tasks instead. As our results indicate, these participants did not show Paap and Noel's pattern. The likely implication is that the success of these criteria when applied to performance in the one-digit-load condition was, in fact, a statistical artifact.

Consistent with this conclusion is the fact that the individuals selected by using these criteria in the a priori tasks tend to be people different from those selected on the basis of performance in the one-digit-load condition of Paap and Noel's (1991) task. That is, application of these criteria in different but comparable tasks does not tend to select a consistent set of people. A crucial assumption underlying Bernstein and Carr's (1996) analysis is that there is an identifiable subset of people who have a particular mental architecture. In dual-route terms, these are the individuals whose assembly route is somewhat faster than their lexical route for low-frequency irregular words. Our results suggest there was no such identifiable subset of individuals in our set of participants (see also Brown et al., 1994) and probably also not in Bernstein and Carr's (1996) set of participants. Rather, whether one is going to have more or less trouble with irregular words appears to depend not on the individual's specific architecture but, as suggested above, mainly on his or her ability to handle the specific words used in that task or condition.

The remainder of the data reported by Bernstein and Carr (1996) also seem to be consistent with this interpretation. Recall that Bernstein and Carr's selection criteria are twofold: Under the one-digit memory load, participants must first show a Frequency × Regularity interaction and must then show a low-frequency regularity effect above the median in size. Most participants selected by the first criterion also are selected by the second criterion, and these individuals have formed the group of interest for the present investigation. There were, however, a few of Bernstein and Carr's participants who met the second criterion but not the first (the participants in Bernstein & Carr's, 1996, Figure 4b and Figure 5b). There were also two other sets of participants in Bernstein and Carr's study, one set that showed a very small low-frequency regularity effect in the one-digit-load condition (Figure 4c and Figure 5c) and one set that showed a reverse low-frequency regularity effect in the one-digit-load condition (Figure 4d and Figure 5d).

Bernstein and Carr (1996) made no particular claims about the nature of these individuals' mental architectures and hence made no claims about what should happen to these individuals' performance in the five-digit-load condition. Nonetheless, a clear pattern emerges. The participants in Bernstein and Carr's Figures 4b and 5b (those who showed large low-frequency regularity effects with a one-digit load but did not show a Frequency × Regularity interaction because they also had a large high-frequency regularity effect) showed a much smaller low-frequency regularity effect with a five-digit memory load in both experiments. That is, their regularity effects decreased from the one- to the five-digit-load conditions, just like the regularity effects for the individuals in the group under investigation. This result is, of course, just what would be expected if their large low-frequency regularity effects were due to seeing most of their difficult irregular words in the one-digit-load condition. On the other hand, the individuals in Bernstein and Carr's Figures 4c and 5c—those who showed a small low-frequency regularity effect in the one-digit-load condition—showed a small increase in the size of their regularity effect in the five-digit-load condition. Finally, the people in Bernstein and Carr's Figures 4d and 5d—those who showed a reverse low-frequency regularity effect in the one-digit-load condition—showed a sizeable increase in the latencies for the low-frequency irregular words in the five-digit-load condition, producing normal or slightly larger than normal low-frequency regularity effects in that condition. These patterns are also what one would expect on the basis of the idea that the truly difficult low-frequency irregular words would be differentially distributed across the one- and five-digit-load conditions for different participants.

The argument that it is the words themselves that are crucial is also potentially consistent with an observation that one can make about situations in which Paap and Noel's (1991) pattern has emerged (e.g., Herdman & Beckett, 1996; Paap & Noel, 1991). In particular, in both situations, the low-frequency regularity effect in the one-digit-load condition was somewhat larger than normal. That is, in most published articles, the size of the regularity effect for naming low-frequency words with no memory load is anywhere from 20 to 35 ms (Andrews, 1982; Brown et al., 1994; Seidenberg et al., 1984; Taraban & McClelland, 1987). In Herdman and Beckett's (1996) data, the effect was somewhat larger than this (44 ms), whereas in Paap and Noel's original study it was inordinately large (over 80 ms).

As noted, it was this fact that caused Pexman and Lupker (1995) to suggest that one source of individual differences might be whether readers experience competition when naming irregular words. That is, some readers may show a regularity effect merely because they have had less exposure to the spelling-to-sound patterns for low-frequency irregular words (possibly the participants in most of the studies in the literature), whereas others may show a larger regularity effect because they also have to cope with a competition situation (e.g., the participants in Paap & Noel's, 1991, experiment). Unfortunately, the present data provided virtually no support for this idea. Thus, at this point, the large low-frequency regularity effects in the one-digit-load condition of Paap and Noel (1991) and Herdman and Beckett (1996) remain unexplained. One possibility that cannot be dismissed out of hand, however, is that these effects may have been a product of item differences. That is, just by chance, it may have been the case that for most of the participants, the low-frequency irregular words in their one-digit-load condition might have been more difficult than the low-frequency irregular words in their five-digit-load condition.

It should also be noted that the rather large low-frequency regularity effects that have been observed by Paap and Noel (1991), by Herdman and Beckett (1996), and for their selected participants, by Bernstein and Carr (1996) in the one-digit-load condition are actually inconsistent with the basic dual-route explanation of Paap and Noel's pattern. That is, the reason that a five-digit-load reduces the size of the regularity effect is because of increased attentional demands. Similarly, a one-digit load would have more attentional demands than no load. Thus, if this type of explanation is correct, the low-frequency regularity effect should actually be smaller, rather than larger, with a one-digit load than with no load.

It is, of course, dangerous to generalize across experiments. However, to our knowledge, no one has yet looked at the effect of a one-digit load versus no load in a controlled way. In fact, our contrast between no-load naming and low-load naming comes closest even though it is also flawed because different words were used in the two tasks and because no-load naming always preceded low-load naming. What this comparison shows, however, is that a one-digit load does appear to require more attention as indexed by the significantly longer response times. What it does not show is that the size of the regularity effect for low-frequency words is altered by a one-digit load. That is, it neither decreased, as would be predicted by Paap and Noel's (1991) dual-route account, nor noticeably increased as occurred in those studies finding Paap and Noel's pattern.

Can Our Failure to Observe Paap and Noel's (1991) Pattern Be Due to Having a Weak Memory-Load Manipulation?

The purpose of the memory-load manipulation in Paap and Noel's (1991) paradigm is to give participants two levels of external attentional demands during naming, one low and one high. The idea is that when maintaining a high load in memory, significantly more attention will be drawn away from the naming task, and thus, the attention-demanding assembly route will be noticeably harmed. Therefore, to be able to answer the question of whether Paap and Noel's pattern can be replicated, one must feel confident that the high-memory-load condition actually does represent a reasonable draw on attention.

To accomplish this, participants in these experiments are told that they should prioritize the memory task. That is, their main goal should be a high level of accuracy in the memory task, and any performance trade-offs that they need to make between the two tasks should not interfere with maintaining this high level of accuracy in the memory task. To try to gauge how well participants have followed these instructions, one can examine accuracy in the memory task. If it is reasonably close to the level of accuracy one would expect if there had been no naming task, one can be confident that participants had been giving the memory task the priority they had been instructed to.

In the version of Paap and Noel's (1991) task reported here (the criterion task) and in our previous experiments with the paradigm (Pexman & Lupker, 1995), memory-task accuracy has been quite high, higher in fact than that observed by Paap and Noel. This is true in spite of the fact that in some of our experiments participants treated the memory task as a speeded-response task, whereas in Paap and Noel's original experiment they did not. This would seem to indicate that our participants were giving the memory task at least as high a priority as Paap and Noel's participants.

On the other hand, two particular results in our experiments might cause some doubt about whether our memory-load manipulation created conditions sufficient for observing Paap and Noel's (1991) pattern in the naming task. First, it was often the case that performance in the memory task was affected by the nature of the word in the naming task (unlike in Paap & Noel's, 1991, experiment). One could argue that this fact, at least partially, compromises our conclusion about the level of priority given to the memory task by our participants. Second, our overall memory-load effects have been small and often nonsignificant. The memory-load effects for low-frequency regular words in these experiments have been especially small. A possible implication could be that in spite of the priority our participants were giving to the memory task, the high-load condition was not attention demanding enough to create the conditions necessary to replicate Paap and Noel's pattern. We believe, however, that there are a number of reasons to dismiss these concerns.

Consider initially the second of these issues, the question of the size of the memory-load effect and the importance of having a large memory-load effect in producing Paap and Noel's (1991) pattern. The first point to make is that across the five experiments reported in Pexman and Lupker's (1995) article and the one experiment reported here, we did in fact find significant memory-load effects twice. In Experiment 1A of Pexman and Lupker (1995), we observed a memory-load effect significant by subjects ($p < .05$) and marginally significant by items ($p < .10$), and in Experiment 2 of Pexman and Lupker (1995; in which a seven-digit load was used in the high-load condition), the memory-load effect was significant by items ($p < .01$) although not by subjects ($.20 > p > .10$). More importantly, in neither of these experiments in which there was evidence of an overall memory-load effect was there any hint that Paap and Noel's pattern was starting to emerge. In fact, quite the opposite was true. In both experiments, the low-frequency regularity effect actually increased from the low- to the high-load condition (by 25 ms in Experiment 1A and by 12 ms in Experiment 2). These increases were, in fact, the two largest increases observed across all six experiments.

Second, beyond the question of statistical significance, the actual means in our experiments suggest that many of our memory-load effects were just as large as Paap and Noel's (1991). Note that Paap and Noel did not observe a significant overall memory load effect either, presumably because they observed a reverse memory-load effect for the low-frequency irregular words (i.e., the three-way interaction). The question then must surely become, if we focus on memory-load effects, what measure can be used to determine whether the attention demands in the high-load condition were sufficient?

According to the attentional-demands assumptions made by Paap and Noel (1991), the main impact of the load is on the assembly route. Thus, one could argue that unless one observes a memory-load effect for the low-frequency regular words (the only set of words that would be badly harmed by a slowed assembly route), one cannot claim that the high memory load is having a sufficient impact.

Unfortunately, the question of whether there actually is a memory-load effect for low-frequency regular words is a major part of the empirical question that is being asked in these replications. That is, the empirical question is whether an increased memory load causes a reduction in the low-frequency regularity effect by causing the latencies for the low-frequency irregular words to decrease, by causing the latencies for the low-frequency regular words to increase, or both. Thus, to use the low-frequency regular word condition as the measure of the strength of the memory-load manipulation creates a circular argument. That is, the claim now boils down to the notion that one has not created sufficient conditions to determine whether Paap and Noel's (1991) pattern can be replicated unless one has already replicated half (or most) of the pattern. As such, we don't believe that the low-frequency regular word condition can be used to provide a reasonable measure of whether our memory-load manipulation was sufficiently strong.

A more reasonable measure would need to focus on data not involving the important effects under investigation. In particular, according to Paap and Noel's (1991) analysis, attentional demands do affect the lexical route as well as the assembly route, which accounts for the latency increase for the high-frequency words under high load. These increases should provide a perfectly acceptable measure for evaluating the strength of memory-load manipulations.

In Paap and Noel's (1991) study, the memory-load effect for the high-frequency regular words appears to have been 12 ms, and the memory-load effect for the high-frequency irregular words appears to have been 18 ms. (The corresponding values reported by Herdman & Beckett, 1996, were 12 ms and 14 ms.) Across our six experiments (in Pexman & Lupker, 1995, and here), memory-load effects for the high-frequency regular words were 12 ms, −6 ms, 18 ms, 25 ms, 16 ms, and 12 ms, and for the high-frequency irregular words our memory-load effects were 15 ms, −3 ms, 21 ms, 36 ms, −20 ms, and 3 ms. Clearly, we have matched or exceeded the size of Paap and Noel's effect (and Herdman & Beckett's, 1996, effect) for the high-frequency regular words five times out of six, and we have exceeded the size of Paap and Noel's effect for the high-frequency irregular words twice (and Herdman & Beckett's, 1996, effect three times). Thus, it appears that according to this criterion, our memory loads acted very similarly to those in experiments that did show Paap and Noel's pattern.

Third, even those individuals in our experiments who showed the strongest memory-load effects (according to our measure) did not show any tendency for the low-frequency regularity effect to decrease under high load. That is, another way of addressing the question of the strength of our memory-load manipulation would be to examine the performance of those participants who clearly did show a memory-load effect for high-frequency words. To that end, the data from the participants in Pexman and Lupker's (1995) Experiments 1A, 1B, and 1C were examined to find the one third of participants who showed the largest memory-load effects for high-frequency words. When these participants' data were analyzed, there was no hint of the crucial Frequency × Regularity × Memory Load interaction ($F_s < 1$ and $F_i < 1$ for both latencies and errors). In fact, for these individuals, the low-frequency regularity effect increased by 15 ms from the low- to the high-load conditions. The fact that those individuals whose lexical routes seem to be very strongly affected by an increased memory load (and, hence, whose assembly routes should be even more strongly affected according to Paap & Noel's, 1991, assumptions) do not show any decrease in the low-frequency regularity effect, clearly suggests that it is not the case that the key to getting the interaction is showing a large memory-load effect.

Fourth, there is one other way we can examine the issue of the strength of the memory-load effects, particularly those for low-frequency regular words, without selecting participants on a part of the task for which we are trying to predict performance. That is, we can consider performance in our a priori tasks and select participants who show slower naming latencies for low-frequency regular words in low-load naming than in no-load naming. Potentially, participants who show a memory-load effect from one task to the other are those whose naming performance is most affected by a digit load and, thus, these might be the participants who would show Paap and Noel's pattern in the criterion task.

To examine this possibility, we selected participants in the upper tertile in terms of a latency increase from no-load naming to low-load naming for low-frequency regular words. These 17 participants showed at least a 74-ms increase in naming latency for low-frequency regular words between no-load naming and low-load naming. The data for these participants in the criterion task are presented in Table 8.

It is interesting that the low-frequency regularity effect did decrease slightly for these individuals. There was, however, no indication of a significant three-way interaction ($F_s < 1$ and $F_i < 1$ for both naming latencies and errors). More importantly, there was no evidence for the other

Table 8
Mean Naming Latencies (in Milliseconds) and Mean Naming Errors (in Percentages) in the Criterion Task for Participants in the Upper Tertile in Terms of a Latency Increase From No-Load Naming to Low-Load Naming for Low-Frequency Regular Words

	Memory load			
	1 digit		5 digits	
Word frequency and regularity	*M*	% error	*M*	% error
Low frequency				
Irregular	651	12.2	659	11.8
Regular	594	0.0	620	1.2
Regularity effect	57	12.2	39	10.6
High frequency				
Irregular	593	0.6	605	1.8
Regular	584	0.0	608	0.0
Regularity effect	9	0.6	−3	1.8

component of Paap and Noel's (1991) pattern, an increased frequency effect for regular words.[4]

Of the four points that we have made in the immediately preceding paragraphs regarding the question of the size of the memory-load effect, Paap and Herdman (1998) have specifically addressed one of them—our claim that the high-frequency word data demonstrate that our participants were being affected by the memory load. Paap and Herdman are unconvinced by these data and, further, go on to suggest that our inability to obtain larger load effects for low-frequency regular words than for high-frequency words is a problem in and of itself, independent of the issues surrounding Paap and Noel's (1991) theorizing. Specifically, Paap and Herdman state that "previous research on attentional effects on word recognition leads to the expectation that a concurrent memory task should interfere more with low-frequency words than with those of high frequency" (p. 846). As such, our failure to obtain a Frequency × Load interaction (at least when considering the regular words) would be inconsistent with the relevant literature.

As Paap and Herdman (1998) acknowledge in their footnote 2, we disagree with their synopsis of that literature. Paap and Herdman cite four articles showing that low-frequency words typically do cause more interference than high-frequency words on performance in a concurrent task (e.g., tone detection; Becker, 1976; Herdman, 1992; Herdman & Dobbs, 1989; Paap & Noel, 1991). The effect of word frequency on concurrent task performance, however, is not the issue here (regardless of whether the concurrent task is primary or secondary). Instead, the issue is what are the effects of a concurrent task on word-recognition performance and, in particular, on performance for high- versus low-frequency words. Regarding that issue, the data in the literature overwhelmingly suggest that the effects for low- and high-frequency words are equivalent (i.e., that there is no Frequency × Attention Load interaction), regardless of whether there is an overall effect of load.

To our knowledge, only the studies involving Paap and Noel's (1991) paradigm have actually looked for this interaction in the situation in which the concurrent task was a memory task. As noted, we have never been able to obtain such an interaction across six experiments, and in fact, our load effects tend to be numerically smaller for the low-frequency regular words than for the high-frequency regular words (although not significantly so). What about the other studies that have used this paradigm (e.g., Bernstein & Carr, 1996; Herdman & Beckett, 1996; Paap & Noel, 1991)? The relevant results are shown in Table 9.

Bernstein and Carr's (1996) Experiment 2 was the most substantial of these experiments, involving 192 participants, two different stimulus sets, and four different types of load. In that experiment, the load effects were 28 ms for the high-frequency regular words and 20 ms for the low-frequency regular words, a nonsignificant difference in the direction opposite from that expected by Paap and Herdman (1998). An inspection of Figure 3 in Bernstein and Carr's article shows that the high-frequency regular word load effect matched or exceeded the low-frequency regular word load effect for three of their four stimulus types. In their Experiment 3, the low-frequency regular words did show a larger load effect than the high-frequency regular words, although this difference was quite small and, again, nonsignificant. There was also no suggestion of this interaction in either Herdman and Beckett's (1996) Experiment 2 or their Experiment 3, although it should be kept in mind that atypical but certainly attention-demanding memory loads were used in these experiments. Even in Herdman and Beckett's Experiment 1, which is cited as a replication of Paap and Noel's (1991) experiment, the difference was small (12 ms for the high-frequency regular words and 28 ms for the low-frequency regular words). More importantly, it grows even smaller if the analysis is done on means (as in the other articles in the literature) rather than on medians.[5] Thus, other than Paap and Noel's experiment, there is no indication in the literature involving Paap and Noel's paradigm that low-frequency regular words are harmed more than high-frequency regular words by a memory load.

What about the other studies mentioned in this context (the studies in which the concurrent task was not a memory task)? In both Herdman's (1992) and Herdman and Dobbs's (1989) experiments, participants in one condition simply made a lexical decision or naming response, whereas in the other condition a "change paradigm" was used in which participants had to forfeit their lexical decision or naming response and respond to a tone if a tone was presented. Having to monitor for tones did cause substantially longer word latencies in this second condition (i.e., 80 ms or more), indicating that participants' attention was being diverted by

[4] A similar analysis was carried out for the participants who were in the upper tertile in terms of their *overall* increase in naming latency from the no-load to the low-load task. As 13 of the 17 selected participants were among the 17 selected on the basis of just their latencies on low-frequency regular word naming, the results of the two analyses were quite similar.

[5] We thank Chris Herdman for doing this analysis and reporting these means to us.

Table 9
Load Effects for Low- and High-Frequency Regular Words in Experiments Involving Paap and Noel's (1991) Paradigm

Experiment	Word type: High frequency regular	Word type: Low frequency regular	Difference
Paap & Noel (1991)			
Experiment 1	12	52	40
Bernstein & Carr (1996)			
Experiment 2	28	20	−8
Experiment 3	11	23	12
Herdman & Beckett (1996)			
Experiment 1 (medians)	12	28	16
Experiment 1 (means)	13	21	8
Experiment 2 (medians)	1	3	2
Experiment 3 (medians)	5	12	7
Mean across experiments other than Paap & Noel's[a]	11	17	6

[a]The data from Herdman & Beckett's (1996) Experiment 1 using means were also not included in these averages.

this additional requirement. The effects, however, were typically a bit larger for the high-frequency words than for the low-frequency words, although never significantly so. Only in Becker's (1976) experiment, when both lexical decisions and tone responses were required sequentially on the same trial, was there evidence that a concurrent task (in this case, an auditory tone-detection task) had a larger effect on low-frequency words than on high-frequency words (the regularity of the words was not reported).

When all of this research is considered, it seems clear that only Becker's (1976) data and Paap and Noel's (1991) data provide any support for Paap and Herdman's (1998) claim that on the basis of previous research, one would expect "that a concurrent memory task should interfere more with low-frequency words than with those of high frequency" (p. 846). Further, given that Becker used a tone-detection task rather than a memory task, the support that those data can provide for this claim are meager at best. In fact, the bulk of the data is quite consistent with the claim that words of low and high frequency are equivalently affected by a concurrent task.

Finally, questions remain concerning the fact that we typically observe effects of the nature of the named word on performance in the memory task. That is, one could argue that if our participants were truly prioritizing the memory task, such effects would not occur. The first point to note is that although Paap and Noel (1991) did not observe effects of this sort, Herdman and Beckett (1996) did. Nonetheless, Herdman and Beckett produced a clear replication of Paap and Noel's pattern. The implication would appear to be that even when one does prioritize the memory task sufficiently, it is possible to obtain these types of effects.

More importantly, to address this issue directly, in Pexman and Lupker (1995) we selected a set of participants from those experiments who clearly did not show any effects of the nature of the words on memory-task performance. These people also did not show a hint of Paap and Noel's (1991) pattern. In fact, the low-frequency regularity effect for these individuals increased by 25 ms from the low- to the high-load conditions.

Our conclusion is that although we readily admit that we have not been able to obtain evidence that our memory-load manipulation creates a memory-load effect for our low-frequency regular words, we do not see how this could explain why we have failed to replicate Paap and Noel's (1991) pattern. By what seems to be a more reasonable measure of the strength of our memory-load manipulation (load effects for high-frequency words), our manipulation seems to be just as effective as those in the studies that did produce Paap and Noel's pattern. Further, even when we consider only those participants who did have large memory-load effects (or those who showed no impact of the nature of the word on their memory-task performance), we find that their data exhibited no greater tendency to show Paap and Noel's pattern than the data of the remainder of the participants. Thus, we see no reason to believe that our failure to show a memory-load effect for the low-frequency regular words can explain our failure to replicate Paap and Noel's pattern in any but the most tautological sense. That is, our failure to obtain a large memory-load effect for our low-frequency regular words is, by definition, at least half of the reason that we do not find a reduced regularity effect in the high-load condition.

Pexman and Lupker's (1995) Account

A secondary goal of this research was to evaluate Pexman and Lupker's (1995) individual differences account, an account based on the idea that those participants who show Paap and Noel's (1991) pattern are the participants who experienced competition when naming irregular words. Succinctly put, this explanation also garnered absolutely no support even though it had more degrees of freedom than Bernstein and Carr's (1996) account. That is, because we had specified no particular criteria for identifying participants, we were able to evaluate a number of criteria in an attempt to find evidence to support our account. None of the criteria provided any indication that they were worthy of further consideration.

Although it might be possible to argue that we simply have not yet provided a full test of our account because we have not yet found good criteria for selecting participants, we find this an unlikely possibility. The main reason is that there appear not to be any real differences between our participants who show Paap and Noel's (1991) pattern and those who do not on the a priori tasks. That is, as our analysis showed, the only difference between these groups in no-load naming and low-load naming was in terms of overall response time. Although this could suggest that overall response time might be a criterion for selecting participants, a tertile split on overall response time in no-load naming and low-load naming also failed to allow us to predict who would show Paap and Noel's pattern.

The implication of these failures seems clear. At least within the set of participants we have at our university, there is no way to predict, a priori, who will show Paap and Noel's (1991) pattern and who will not. Neither Bernstein and Carr's (1996) account nor Pexman and Lupker's (1995) account are viable individual differences explanations for the failures to replicate Paap and Noel's pattern. We readily acknowledge the possibility that at least some participants at places where the pattern has been obtained might actually be qualitatively different from the participants at the University of Western Ontario and that there may be some way to identify these people a priori. At present, however, we have neither what we would consider a viable explanation of why that might be true nor what we would consider a workable idea of how it could be accomplished.

References

Andrews, S. (1982). Phonological recoding: Is the regularity effect consistent? *Memory & Cognition, 10,* 565–575.

Baron, J., & Strawson, C. (1976). Use of orthographic and word-specific knowledge in reading words aloud. *Journal of Experimental Psychology: Human Perception and Performance, 2,* 386–393.

Becker, C. A. (1976). Allocation of attention during visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 2,* 556–566.

Bernstein, S. E., & Carr, T. H. (1996). Dual-route theories of spelling to pronunciation: What can be learned from concurrent task performance? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,*86–116.

Brown, P., & Besner, D. (1987). The assembly of phonology in oral reading: A new model. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 471–490). Hillsdale, NJ: Erlbaum.

Brown, P., Lupker, S. J., & Colombo, L. (1994). Interacting sources of information in word naming: A study of individual differences. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 537–554.

Coltheart, M. (1978). Lexical access in simple reading tasks. In G. Underwood (Ed.), *Strategies of information processing* (pp. 151–216). New York: Academic Press.

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. (1991). The effect of spelling-to-sound regularity on naming in French. *Psychological Research/Psychologische Forschung, 53,* 3–12.

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior, 12,* 627–635.

Herdman, C. M. (1992). Attentional resource demands of visual word recognition in naming and lexical decisions. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 460–470.

Herdman, C. M., & Beckett, B. (1996). Code-specific processes in word naming: Evidence supporting a dual-route model of word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 1149–1165.

Herdman, C. M., Beckett, B., & Stolpmann, K. (1993, July). *Code-specific interference in phonological recoding: Evidence supporting a dual-route approach.* Paper presented at the meeting of the Canadian Society for Brain, Behaviour, and Cognitive Science, Toronto, Ontario, Canada.

Herdman, C. M., & Dobbs, A. R. (1989). Attentional demands of visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 124–132.

Jared, D. (1995, November). *Spelling–sound consistency affects the naming of high-frequency words.* Poster presented at the 36th Annual Meeting of the Psychonomic Society, Los Angeles, CA.

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

Lupker, S. J., Brown, P., & Colombo, L. (1997). Strategic control in a naming task: Changing routes or changing deadlines? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 570–590.

Monsell, S., Patterson, K. E., Graham, A., Hughes, C. H., & Milroy, R. (1992). Lexical and sublexical translation of spelling to sound: Strategic anticipation of lexical status. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 452–467.

Paap, K. R., & Herdman, C. M. (1998). Highly skilled participants and failures to redirect attention: Two plausible reasons for failing to replicate Paap and Noel's effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 845–861.

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

Pexman, P. M., & Lupker, S. J. (1995). Effects of memory load in a word naming task: Five failures to replicate. *Memory & Cognition, 23,* 581–595.

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

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

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

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

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.

Appendix

Experimental Stimuli

Low frequency		High frequency	
Irregular	Regular	Irregular	Regular
No-load naming stimuli			
bowl	beam	are	big
brow	bunt	bear	block
choose	cub	blind	broke
deaf	deed	broad	got
flood	float	child	group
fuse	gate	do	held
monk	hunt	enough	help
pint	mode	gone	less
plow	peel	great	out
shoe	pump	lose	place
soot	slat	love	sit
steak	stab	ones	still
swamp	wail	phase	take
wan	wake	put	thin
wash	wane	some	will
Low-load naming stimuli			
bush	bean	blood	did
cease	carve	break	each
doll	dusk	does	fact
flown	fade	front	him
hind	grape	gross	life
host	lent	mind	main
mow	mill	none	name
pear	plank	post	not
plaid	plump	pull	page
rouse	rink	prove	see
smooth	sank	shall	stop
spook	slam	to	tell
sweat	swore	watch	turn
tomb	wig	what	when
wad	yell	wood	while
Criterion task (Paap & Noel, 1991, Experiment 1) stimuli			
bury	buds	been	best
caste	canes	both	book
comb	coil	come	came
crow	curl	done	dark
glove	grade	door	deep
lure	lump	foot	flat
lute	lode	give	game
pour	pops	good	gain
ruse	rump	have	high
sew	sock	most	more
sans	sage	move	miss
sues	suck	said	same
sown	sobs	says	seem
wand	wade	sure	soon
warn	weed	touch	train
warp	wick	want	wall
wasp	weld	warm	wage
wily	wilt	were	well
wool	woke	word	west
worm	wink	work	week

Received January 31, 1996
Revision received August 23, 1996
Accepted August 31, 1996 ■