This is to certify that the

dissertation entitled

HOW DO READERS PRONOUNCE IRREGULAR WORDS IN ENGLISH? A COMPARISON OF SINGLE AND DUAL ROUTE FUNCTIONAL ARCHITECTURES

presented by

STUART E. BERNSTEIN

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Psychology

T.H. Carr
Major professor

Thomas H. Carr

Date 2/29/96

HOW DO READERS PRONOUNCE IRREGULAR WORDS IN ENGLISH?
A COMPARISON OF SINGLE AND DUAL ROUTE FUNCTIONAL ARCHITECTURES.

By

Stuart E. Bernstein

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Department of Psychology

1996

ABSTRACT

HOW DO READERS PRONOUNCE IRREGULAR WORDS IN ENGLISH? A COMPARISON OF SINGLE AND DUAL ROUTE FUNCTIONAL ARCHITECTURES.

By

Stuart E. Bernstein

Paap and Noel (1991) found that in naming words aloud a concurrently performed memory task diverted the attention required to assemble a pronunciation, causing a decrease in naming latencies for low frequency words with irregular spellings and an increase in naming latencies for other word types. The current experiments showed the importance of two factors in replicating their results: (1) Individual differences in susceptibility to the frequency by regularity interaction in naming and (2) The way in which spelling-to-sound regularity is operationally defined. Experiment 1 established that an indicator of individual differences, the frequency by regularity interaction, was not significantly altered by the dual task paradigm. This indicator was used to classify subjects in Experiments 2 and 3. In Experiment 2, spelling to sound regularity was defined by grapheme-to-phoneme correspondences (GPCs). For subjects who were relatively sensitive to GPC regularity an increase in concurrent task demands reduced the interference between an incorrect GPC assembled pronunciation and a correct retrieved pronunciation. In Experiment 3, for subjects who were relatively insensitive to GPC regularity but were sensitive to neighborhood consistency, an increase in concurrent task demands reduced the interference that an inconsistent word suffered from a higher frequency enemy (a word with a similar spelling but different pronunciation). Taken together, these results indicate that different subjects experience different amounts of interference when naming different stimuli. The performance of GPC sensitive subjects on GPC regular/irregular stimuli is best explained by a model with a dual route architecture while the performance of consistency sensitive subjects on other neighborhood consistent/inconsistent stimuli is best explained by a model with a single route architecture.

Acknowledgments

I would like to thank Tom Carr, Rose Zacks, Fernanda Ferreira, Lauren Harris, and Brad Rakerd, for their guidance. I would also like to thank Rick DeShon for his statistical advice.

TABLE OF CONTENTS

LIST OF TABLES v

LIST OF FIGURES vi

PAGE

1 Introduction 1

2 A Critical Analysis of Post-Hoc Subject Classification 30

3 Experiment 1 40

4 Experiment 2 58

5 Experiment 3 69

6 General Discussion 81

APPENDICES

A Measures of Print Exposure 90

B Experiment 1 Naming Stimuli 94

C Experiment 2 Naming Stimuli 95

D Experiment 3 Naming Stimuli 96

E Split Half Analysis 97

LIST OF REFERENCES 99

LIST OF TABLES

NUMBER PAGE

1 Experiment 1: Proportion of errors in naming. 48

2 Results of the split-half reliability analysis 49

3 Results of the split-half reliability analysis for categorical measures 50

4 Pearson correlation matrix for single and dual task naming performance 51

5 Pearson correlation matrix for experience measures and single task naming performance 52

6 Pearson correlation matrix for experience measures and single task naming performance 53

7 Experiment 2: Naming & Memory Error Rates 63

8 Experiment 3: Naming & Memory Error Rates 76

LIST OF FIGURES

NUMBER		PAGE
1	The time course of processing in a dual route model	6
2	Distribution of low frequency regularity effects	33
3	Frequency * Regularity interaction for single vs. dual task naming	47
4	Experiment 2 naming latencies	62
5	Effect of Individual Differences in Sensitivity to GPC Regularity	66
6	Experiment 3 naming latencies	74
7	Effect of Individual Differences in Sensitivity to GPC Regularity	78

CHAPTER 1

Introduction

A fluent reader of English is capable of correctly pronouncing familiar and relatively unfamiliar words as well as generating pronunciations for novel letter strings. This performance could be achieved by taking advantage of the relationship between individual written units, or graphemes, and individual sound units, or phonemes. The GPC hypothesis is the idea that grapheme to phoneme correspondences of the sort described by linguists (Venezky, 1970; Wijk, 1969) are the units of analysis in spelling-to-sound recoding. This hypothesis is supported by findings that naming latencies are sensitive to whether or not a word follows GPC rules (Baron & Strawson, 1976; Gough & Cosky, 1977; Stanovich & Bauer, 1978). A recent test of the GPC hypothesis using a developmental computer model (Coltheart, Curtis, Atkins, & Haller, 1993) resulted in correct pronunciation of 78% of the words in Seidenberg and McClelland's (1989) 2,897 word corpus. Coltheart and colleagues interpret this as confirmation that a GPC rule based system is capable of pronouncing much of what would be a reader's core vocabulary, assuming that the corpus of words they chose is representative.

The GPC hypothesis has limitations, which are explained in the following paragraphs. The consequences of these limitations are that adopting the GPC hypothesis forces one to accept additional hypotheses to overcome its limitations. These hypotheses are the independent processes hypothesis and the relative finishing times hypothesis (often over-simplified as the delayed phonology hypothesis). Together with the GPC hypothesis, these three hypotheses make up dual route model of spelling-to-sound. In dual route models, an assembled phonology mechanism is paired with a functionally independent retrieved phonology mechanism (Baron & Strawson, 1976; Coltheart, 1978; Coltheart, et al., 1993; Monsell, Patterson, Graham, Hughes, & Milroy, 1992; Paap, McDonald, Schvaneveldt, & Noel, 1987). All three hypotheses comprising dual route models have been challenged (see Humphreys & Evett, 1985; Van Orden, Pennington,

& Stone, 1990 for reviews) but not decisively (for a response see Coltheart et al., 1993; Paap, Noel, and Johansen, 1992). This paper will review the evidence about all 3 hypotheses and present the results of 3 experiments intended to explore the predictions of the GPC hypothesis and its alternatives.

Limitations of the GPC Hypothesis

A GPC system is limited in the spelling-to-sound relationships that it can capture. This limitation is due to the fact that GPC rules incorporate a minimal amount of context. For example, in Venezky's (1970) rules as well as in the Coltheart et al. (1990) simulation model, the context sensitivity of a grapheme to phoneme correspondence is limited to a single grapheme neighbor. This neighbor does not need to be adjacent to the grapheme in question, which allows relationships such as the change in vowel sounds in words with final "e" to be captured. The benefit of restricting context sensitivity is the ability to generalize. That is, a limited amount of knowledge can be applied in a wide variety of contexts. The drawback to limited context sensitivity is that a GPC rule-based system cannot explain how readers are able to correctly pronounce words which are exceptions to the rules. For example in the word "pint," the grapheme "I" followed by a word final consonant cluster corresponds to /ai/ rather than /I/, the latter of which is more frequent in this context. Similarly, in the word "caste" the pattern "a_e" corresponds to /æ/ rather than to /e/, the latter being more frequent in this context. Coltheart and colleagues' (1993) test of the GPC hypothesis confirmed this limitation. The model was not capable of correctly pronouncing all words in the corpus that it was trained on -- it mispronounced 22%. According to Coltheart, these words were GPC exception words. Because GPC rules cannot correctly generate pronunciations for exception words, any model of pronunciation that uses GPC assembly must employ a separate mechanism to capture those relationships between spelling and sound that are not regular grapheme-to-phoneme correspondences.

The Independent Processes Hypothesis

The independent processes hypothesis is the claim that the retrieved and assembled phonology mechanisms in a dual route model are functionally independent. That is, the processing of information in one mechanism does not affect the other. The independent processes hypothesis is supported by studies of clinical cases of acquired and developmental dyslexia. Typically, a double dissociation occurs in deficits in the retrieval and assembly of pronunciations such that acquired dyslexias can be classified as falling into one of two categories (McCarthy & Warrington, 1990; Patterson & Coltheart, 1987). Developmental dyslexias have also been classified as displaying a dissociation similar to that observed in acquired dyslexias. Patients suffering from surface dyslexia are capable of reading regular words aloud but have trouble with irregular forms (Marshall & Newcombe, 1973; McCarthy & Warrington, 1986; Patterson, Marshall, & Coltheart, 1985). In developmental surface dyslexia, children have difficulty acquiring retrieved phonology skills (Coltheart, Masterson, Byng, Prior, & Riddoch, 1983). This supports the independent processes hypothesis because a dual route model could produce this behavior if the retrieved phonology mechanism were damaged. Patients suffering from phonological dyslexia can correctly pronounce regular and irregular words but perform poorly on nonwords and rare words (Beauvois & Derousene, 1979; Funnell, 1983). In developmental phonological dyslexia, children have difficulty assembling pronunciations (Temple & Marshall, 1983). This supports the independent processes hypothesis because a dual route model could produce this behavior if the assembled phonology mechanism were damaged.

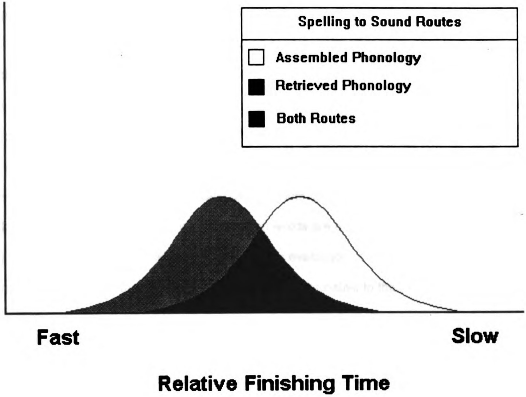
While alternative explanations of the dyslexia data are conceivable (e.g., Humphreys & Evett, 1985), approaches other than the independent processes hypotheses have not yet been successful (see Coltheart et al., 1993 for a critical review). For example, Patterson, Seidenberg, and McClelland (1990) tested the possibility that damage to Seidenberg and McClelland's (1989) single route PDP assembled phonology network could result in performance similar to that displayed in surface dyslexia. By removing hidden units, they were able to get the model to regularize proportions of exception words similar to human patients. However, when the model was damaged in this fashion, it performed poorly in pronouncing nonwords, unlike human patients with surface dyslexia, who pronounce nonwords very well.

This problem with the performance of the single route PDP model has resulted in one recent change from a single route PDP model that rejects the independent processes hypothesis (Seidenberg & McClelland, 1989; Plaut & McClelland, 1993) to a dual route PDP model that incorporates the independent processes hypothesis (Plaut et al., 1994). With the adoption of a dual route architecture, the dual route PDP model is structurally similar to some implementations of the dual route GPC model (e.g. Monsell et al., 1992). However, this similarity is limited to the independent processes hypothesis. Assembled phonology functions quite differently in PDP models, which unlike traditional dual route models do not adopt the GPC hypothesis.

The Time Course of Processing

The second limitation of GPC rules is their inability to account for the frequency by regularity interaction in naming. The form of this interaction is that the impact of spelling-to-sound regularity is restricted to low frequency words -- naming latencies for words that are high in frequency have shown little impact of spelling-to-sound regularity (e.g. Seidenberg, Waters, Barnes, & Tannenhaus, 1984; Taraban & McClelland, 1987; but see Jared, 1995 for a counter-example). In a dual route architecture this performance is explained by a relatively complex set of assumptions about the time course of processing in the two routes from spelling to sound. A summary of these time course assumptions follows (see Paap et al., 1992 for a full discussion). Figure 1 displays two distributions that are intended to represent the time taken for retrieved and assembled phonology each to deliver the pronunciation for all words that they might encounter. Processing time, which is represented along the abcissa, increases from relatively short to relatively long. Frequency sensitivity causes variation in the finishing time of the retrieved phonology mechanism such that pronunciations of higher frequency words are generally retrieved quickly. The finishing time of assembled phonology is determined by the frequency with which the grapheme to phoneme correspondences within a word have been encountered in the past. In a dual route model, GPC regularity influences response times when the phonology that is assembled for a string of letters does not match the phonology that is retrieved for that string. This can only happen when the two routes deliver their answers concurrently, represented as the darkly shaded area of Figure 1. Low frequency words tend to evoke responses that fall within this region because their pronunciations are retrieved relatively slowly. High frequency words tend to evoke a fast response from retrieved phonology, represented by the lightly shaded area of Figure 1.

Figure 1: Time course of processing in a dual route model



Van Orden and colleagues (1990) simplify this relatively complex set of assumptions about the time course of processing into a single hypothesis called delayed phonology. "The delayed phonology hypothesis assumes that phonologic codes are late sources of constraint in lexical coding relative to direct access from orthographic codes" (Van Orden et al., 1990, p. 490). The observable consequences of this hypothesis on performance are that for skilled readers the phonological characteristics of words should have a relatively small impact on word identification latencies. The dual route model makes this prediction because skilled readers should display increased reliance on retrieval as opposed to assembly in the pronunciation of words. It is important to note that the delayed phonology hypothesis is an oversimplification of the time course assumptions outlined in the previous paragraph. The delayed phonology hypothesis is only an accurate representation of the dual route model's prediction for high frequency words. For low frequency words, phonological codes generally do contribute to responses. This is true even among skilled readers, the only difference being which words are actually low in frequency for a particular subject. This distinction will be important in the evaluation of the evidence against the time course predictions in dual route models. Van Orden's alternative to the delayed phonology hypothesis is the phonological mediation hypothesis, which states that phonological codes obligatorily contribute to the recognition of all words for all readers.

The lexical decision task has been used to evaluate the phonological mediation hypothesis (Coltheart, Davelaar, Jonasson,& Besner, 1977; Rubenstein, Lewis, & Rubenstein, 1971). Rubenstein et al. (1971) presented subjects with two types of word stimuli, homophones such as "SALE" and "WEAK" and non-homophones such as "BATH" and "PINK". They found that acceptance latencies to homophones were longer than acceptance latencies for non-homophones. Coltheart et al. (1977) did not find this effect in their replication attempt. Rubenstein et al. interpreted this finding to mean that phonological codes are used to activate lexical entries, and the more lexical entries that are activated, the longer lexical decision takes. Subjects were also presented with two types of nonword stimuli, pseudohomophones like "BURD" and "GROE" and nonpseudohomophones such as "ROLT" and "SAFT". Both Rubenstein and

Coltheart found that rejection latencies to pseudohomophones were longer than rejection latencies for nonpseudohomophones. Rubenstein et al. interpreted their findings as compatible with the view that phonological codes are contributing to responses, activating the lexical entry for the pseudohomophones' real word counterparts, e.g. BIRD and GROW. Coltheart et al. (1977) questioned this interpretation because it is based on the negative responses, which are typically slower than positive responses in lexical decision. Therefore, effects of homophony on negative responses could be post-access -- they do not necessarily reflect the processes involved in the recognition of real words.

To avoid the problems with the lexical decision task, Van Orden and colleagues (Van Orden, 1987; Van Orden, Johnston, & Hale, 1988) tested the delayed phonology hypothesis using rapid semantic categorization. They found that subjects produced more false positive responses to a word that sounded like a category exemplar such as "ROWS" for the category "A FLOWER" than for a control item that was spelled similarly such as "ROBS." Van Orden (1987, Experiment 3) demonstrated that the elevated false positive rate to pseudohomophones was independent of base word frequency. Van Orden argued that this effect was not likely to be post-access because the elevated false positive rate was on the relatively fast "yes" responses in semantic categorization. This contradicts the delayed phonology hypothesis, which does not predict a large influence of phonology on a quickly generated response. The elevated false positives also appeared in an experiment where the target was masked after a brief presentation (Van Orden, 1987, Experiment 2), a manipulation that Van Orden argued should make responses sensitive to the earliest source of constraints on encoding. Heightened false positives to homophones in this masked condition suggests that these early constraints are phonological. Comparison of the distributions of latencies in semantic categorization revealed that false-positive latencies to homophones like "ROWS" were highly overlapping with true positive responses for targets like "ROSE" (Van Orden et al., 1988). If the output of a phonological encoding mechanism were delayed, as dual route theory predicts, the false positive responses to homophone targets would have been relatively slower. Based on this evidence from semantic categorization Van Orden et al. (1990) reject the delayed phonology hypothesis in favor of the the phonological mediation hypothesis, according to which a primary factor in word recognition is phonology.

A problem with Van Orden's homophone experiments was discovered in a replication attempt by Jared and Seidenberg (1991). Jared and Seidenberg suspected that part of the semantic priming effects that Van Orden and colleagues observed could be due to conscious expectations, which would be facilitated by predictable targets. Van Orden's stimuli included many categories with small numbers of exemplars, which would facilitate conscious predictions. Jared and Seidenberg (Experiment 2) used broader categories to reduce the predictability of the targets and found that increased false positive responses were restricted to low frequency targets. This result is consistent with the time course predictions of dual route theories.

The predictions of delayed phonology have also been explored through the use of pseudohomophones as primes and targets in an associatively primed naming task. The delayed phonology hypothesis predicts that a pseudohomophone like "tayble" should not prime a target like "chair" at short SOAs because phonological processing of "tayble" would not be completed early enough to affect the processing of "chair." Lukatela and Turvey (1991) found that naming a word like "chair" was facilitated by the prior presentation of a pseudohomophone of a semantic associate like "tayble" (Experiments 1 & 2). Furthermore, they found that this priming was of equal magnitude at both short (280 ms) and long (500 ms) SOAs (Experiment 4). Lukatela and Turvey concluded that the delayed phonology hypothesis was disproven. However, while Lukatela and Turvey did manipulate SOA to influence the time course of processing, they ignored target frequency, which also affects the time course of processing. "Tayble-chair" priming only disproves the time course assumptions in dual route theories if it occurs for both low frequency targets and high frequency targets. Since Lukatela and Turvey did not account for the frequency of their targets, their conclusion of rejecting the time course predictions in dual route theories is unjustified or at least premature.

Much of the research on delayed phonology has been done using monosyllabic words. Henderson, Dixon, Petersen, Twilley, and Ferreira (1995) tested the delayed phonology hypothesis using both monosyllabic and polysyllabic words in a transsaccadic word recognition task. In this task, subjects fixated on a cross while a preview string was presented extrafoveally. Subjects were instructed to execute an eye movement to the string when they detected it and to make a lexical decision. The benefit of pre-exposure to a target was measured. Because the preview string was only viewed peripherally and for a brief duration, Henderson et al. argued that any effect of the preview string on target processing can be attributed to the influence of information from the initial stages of lexical processing. They found that when the preview and target strings were identical polysyllabic words, there was a greater preview benefit for words with phonologically regular vs. phonologically irregular initial trigrams. There was no effect of regularity when the stimuli were monosyllabic words. In a post-hoc analysis, Henderson et al. found that the preview benefit for phonologically regular polysyllabic words was greater for low frequency words than for high frequency words. The finding that phonological codes are more important for the recognition of low frequency words is consistent with the delayed phonology hypothesis in dual route models.

Flexible Coding

Some versions of the dual route model of pronunciation (e.g. Pugh, Rexer, & Katz, 1994; Monsell et al., 1992) and larger models of word recognition and pronunciation that incorporate a dual route approach to pronunciation, such as the parallel coding systems model (Carr & Pollatsek, 1985), include an operational principle allowing variation in the extent to which phonology is employed in recognition. Pugh et al. (1994) refer to this as flexible coding. These models predict that subjects will suspend the use of phonology when it is made detrimental through task demands, and that subjects will rely more heavily on phonology when it is beneficial to performance. Because this could change the relative time course of processing in the two routes, the principle of flexible coding has been criticized because it makes these time course assumptions difficult to falsify. Nevertheless, the two issues are linked and evaluation of the time course assumptions should also consider flexible coding.

Flexible coding has been tested in the lexical decision task through the use of pseudohomophone distractors. This manipulation should make the use of phonological recoding detrimental to performance because it would lead to false positive responses to pseudohomophonic nonword distractors, strings like "bote" and "tite." Rubenstein, Lewis, and Rubenstein (1971) found that using pseudohomophones as nonwords in lexical decision increased rejection latencies. This suggests that subjects were phonologically recoding, even though it is detrimental to performance -- direct evidence against flexible coding. As mentioned earlier, this experiment has been criticized because the effect occurs on rejection latencies, which do not necessarily reflect encoding processes. Consequently, research that followed looked for effects on "yes" responses in lexical decision.

Pugh et al. (1994) solved this problem by comparing the performance on the "yes" trials of a lexical decision experiment for groups of subjects who had either no pseudohomophones in their nonwords or a high proportion of pseudohomophones. The flexible coding hypothesis can be interpreted as predicting that if subjects can shut down phonological recoding, both rejection and acceptance latencies in lexical decision should be influenced. Pugh et al. found that responses to the word targets in the pseudohomophone condition were faster than in the no pseudohomophone condition, just as accurate, and less sensitive to spelling-to-sound regularity. They concluded that subjects who expected to see pseudohomophone distractors protected themselves from this detriment by relying less heavily on phonological recoding -- evidence for flexible coding. This resulted in faster responses for subjects in the pseudohomophone condition because they relied more heavily on the retrieved phonology route, which is itself faster than the assembly route.

Unlike the lexical decision data, evidence from semantic categorization tasks does not support flexible coding. Jared and Seidenberg (1990) tested the flexible coding hypothesis using the rapid semantic categorization task. In Experiments 2 and 3 they manipulated the proportion of homophones in their stimulus list and observed the effects on false positive responses to judgments such as a "ROWS" is a "FLOWER." Flexible coding predicts that subjects will suspend the use of phonological recoding and make fewer false positive errors when they expect homophones. They found that false positive error rate was not affected by the proportion of homophones in a list. These results do not support flexible coding.

Data from the naming task support flexible coding. In naming, the use of phonological encoding when exception words are targets is logically detrimental to performance, as it would lead to mispronunciations. If flexible coding is possible, subjects should protect themselves from this detriment by inhibiting the use of phonology when they expect to see exception words. Monsell et al. (1992) compared performance in a condition where subjects were required to use assembled phonology, naming lists of nonwords mixed with exception words, with a condition where assembled phonology was not required, naming lists of pure exception words. They found that subjects made significantly fewer regularization errors to exception word targets when they were named in pure lists versus lists in which they were mixed with nonwords. This supports the hypothesis that subjects can suspend the use of phonological information when it is detrimental to performance. Monsell et al. (1992) also tested the possibility that retrieval rather than assembly could be flexibly employed. They found that subjects were able to name nonwords faster when they appeared in pure blocks versus blocks in which they were mixed with words (Experiment 1), but only when these words were low in frequency (Experiment 2). Naming latencies for nonwords mixed with high frequency words were not significantly different than the pure block of nonwords control condition (Experiment 2). According to flexible coding this happened because subjects who were expecting pure lists of nonwords did not need to wait for confirmation of their assembled pronunciations from the addressed phonology mechanism. Monsell and colleagues reasoned that if subjects adopted this strategy, then they should tend to regularize an unexpected low frequency exception word that appears in a list of nonwords, which subjects did do. This evidence supports the hypothesis that subjects will ignore the output of addressed phonology when naming lists consisting predominantly of nonwords.

The predictions of flexible coding are also supported by experiments using tachistoscopic recognition (Hawkins, Reicher, Rogers, & Peterson, 1976; Carr, Davidson, & Hawkins, 1978; see

Carr & Pollatsek, 1985 for a review). Hawkins et al. (1978) presented subjects with a masked word (e.g. sent) followed by a forced choice judgment between it and a homophonic alternative (e.g. sent /cent). Subjects were less accurate at homophonic response alternatives than non-homophonic response alternatives, but only when there were few homophonic pairs in the list -- which suggested to Hawkins et al. that subjects relied less heavily on phonological encoding when it was detrimental to performance, as in the lists with high proportions of pseudohomophones. Carr et al. (1978) found that the pseudoword superiority effect in tachistoscopic recognition was also subject to expectancy effects. The pseudoword superiority effect is the finding that subjects are more accurate at detecting letters in pseudoword strings than nonword strings (Baron & Thurstone, 1973). Carr et al. (1978) manipulated the extent to which subjects needed to rely on phonological encoding by varying the proportion of pseudowords and nonwords in a list. They found that when subjects expected to see both words and pseudowords, they displayed a word and pseudoword superiority effect. Subjects were not able to identify letters in pseudowords any better than in nonsense strings when they expected either purely words or purely nonsense strings. Carr et al. interpreted this to mean that subjects could vary the extent to which they relied on phonological encoding -- evidence for flexible coding. But their evidence also suggested that this flexibility was hard to induce. This may explain the mixed results in other tasks that have been used to evaluate flexible coding.

Conclusions About Dual Route Time Course Assumptions and Flexible Coding

Two conclusions are possible based upon the research reviewed above. First is that contrary to some existing reviews (Humphreys & Evett, 1985; Van Orden et al., 1990), the time course assumptions of dual route models are not falsified by research in naming, lexical decision, semantic categorization, and tachistoscopic recognition. The second conclusion is that subjects may be able to flexibly employ different recoding strategies. While this evidence is consistent with a dual route architecture of some type, it does not constitute decisive evidence in favor of dual route GPC models because there are alternatives to the GPC hypothesis. These alternatives are more parsimonious than dual route GPC models in that they do not face the limitations which require additional assumptions about independent processes and time course of processing.

Alternatives to the GPC Hypothesis

An alternative solution to the problems with GPC based assembled phonology is to abandon the notion of grapheme to phoneme correspondences and use larger functional spelling units. Proposed units range in size from the onsets and rimes of syllables (Bowey, 1990; Treiman & Chafetz, 1987; Treiman & Zukowski, 1988; Treiman & Zukowski, 1990) to whole words (Brooks & Miller, 1979). Multiple size approaches have also been proposed (Glushko, 1979; Glushko, 1981) as well as PDP models that in theory do not explicitly adopt any single level of analysis (Van Orden et al., 1990).[1] The benefit of using larger units is a gain in context sensitivity, which allows spelling-to-sound relationships to be represented more specifically. A system with greater context sensitivity can correctly pronounce a greater proportion of words than a system with restricted context sensitivity. A model that uses units of analysis larger than GPCs is therefore capable of doing so in a single route architecture. This approach is taken by analogy models (Brooks & Miller, 1979; Glushko, 1979; 1981), and PDP models (Lukatela, Turvey, Feldman, Carello, & Katz, 1989; Seidenberg & McClelland, 1989; Plaut & McClelland, 1993; Seidenberg, Plaut, McClelland, & McRae, 1994; Van Orden et al., 1990). While these models are functionally diverse, they share at least 1 feature. In all of them spelling-to-sound relationships are not described in terms of GPC regularity but rather in terms of orthographic neighborhood consistency.

Neighborhood Consistency

The orthographic neighborhood of a monosyllabic word is the set of other words that can be produced by combining other onsets with its rime. For example, the word "mint" consists of

---

[1] PDP models must adopt a representational scheme for input representations, which can be construed as a unit of analysis. However, Van Orden and colleagues claim that their choice of input representations was an arbitrary, and the performance of the model is independent of the particular representational scheme chosen. This claim remains unproven.

the onset "m" and the rime "int." The orthographic neighbors of "mint" are "lint", "hint" , "tint", and "pint." The orthographic neighborhood of a word can be either consistent or inconsistent. Words in consistent neighborhoods have their bodies pronounced similarly across all onsets, whereas words in inconsistent neighborhoods have their bodies pronounced differently with some onsets than with others. The naming of a word in a consistent neighborhood should be facilitated by its neighbors. For example, the naming of "halt" should be facilitated by its friends, the sound alike neighbors "malt" and "salt". The naming of a word in an inconsistent neighborhood should be inhibited by its enemies, the different sounding neighbors. For example, the naming of "pint" should be inhibited because it does not sound like its neighbors "mint", "lint", and "hint".

Glushko (1979) was among the first to test the idea that the consistency of a word's orthographic neighborhood, rather than whether or not a word follows GPC rules, should affect naming latencies. He found that words from inconsistent neighborhoods took longer to name than words from consistent neighborhoods, even among words classified as "regular" by a GPC scheme. This finding cannot be accommodated by a dual route GPC theory, which predicts that consistency should be irrelevant to the naming of any regular word. Glushko also found that nonwords which are spelled similarly to words with inconsistent bodies took longer to name than nonwords spelled similarly to words with consistent bodies. He interpreted these findings as supporting the notion that words and nonwords alike are pronounced by analogy to a set of candidate lexical representations that are activated by any orthographic input. This finding has been interpreted as falsifying dual route models, based on the assumption that they separate the impact of frequency and GPC regularity. However, this is a problem for a dual route GPC model only if the rules are not allowed to vary in strength (for a discussion see Paap et al., 1992). The Coltheart et al. (1993) simulation model with rule strength is able to accommodate these findings.

Glushko's paper had a lasting impact on the field in that the majority of new models of visual word recognition that followed the publication of his paper depicted spelling-to-sound relationships as neighborhood consistency rather than GPC regularity based. Jared et al. (1990) note that the lasting impact of this paper is surprising, given further investigations showing that the effects of inconsistency were due to pronouncing conflicting neighbors within the same block of trials (Seidenberg, et al., 1984). Subsequent efforts to replicate consistency effects have met with mixed results (for a review see Jared, McRae, & Seidenberg, 1990; Patterson & Coltheart, 1987). Consistency effects are not generally found among high frequency words (Andrews, 1982; Kay & Bishop, 1987; Seidenberg et al., 1984; Stanhope & Parkin, 1987; Taraban & McClelland, 1987; but see Jared, 1995, who does find consistency effects among high frequency words). This result alone is not a problem, since analogy and PDP accounts of naming predict that consistency effects will interact with frequency such that consistency effects are relatively weak among high frequency words. However, these models also predict strong consistency effects among low frequency words, which have not always been found (Seidenberg et al., 1984, Experiment 4; Andrews, 1982; Stanhope & Parkin, 1987, Experiment 2; Taraban & McClelland, 1987).

Patterson and Coltheart (1987) interpret these mixed findings as an indication that GPC regularity, not neighborhood consistency, is relevant to the naming of words. Jared et al. (1990) offer an alternative explanation for the unreliability of consistency effects. They suggest that there is a problem in how consistency has been measured. Consistency is most often defined by a <u>type</u> based measure, <u>neighborhood size</u>, the number of word forms with a particular spelling, and pronunciation regardless of the frequency of these words (e.g. Glushko, 1979, Experiment 1). In this sort of measure, an inconsistent word is one with many orthographically similar but phonologically distinct neighbors. An alternative way to compute consistency is a <u>token</u> based measure such as <u>neighborhood frequency</u>, which considers the relative frequency of a pronunciation within its neighborhood. In this sort of measure, an inconsistent word is one that has orthographically similar but phonologically different neighbor(s) that are higher in frequency than it is. In a meta-analysis of naming experiments that used neighborhood consistency as an independent variable, Jared et al. found that failures to find effects of consistency were correlated with failures to control for neighborhood frequency.

Jared et al.'s own experiments compared neighborhood size and neighborhood consistency. In their first two experiments, they explored the impact of <u>neighborhood frequency</u>

and found that subjects named words with high frequency enemies significantly more slowly than words with low frequency enemies (Experiment 1) and that the frequency of the enemies of a word interacted with the frequency of the friends (Experiment 2) such that the impact of enemy frequency was greatest for words with low frequency friends. In their third experiment, Jared and colleagues examined the impact of <u>neighborhood size</u>, the number of friends and enemies a word has rather than their average frequency. Unlike Experiment 2, where neighborhood type (friends vs. enemies) interacted with consistency, in Experiment 3 they found no interaction between consistency and neighborhood type. This suggests that consistency effects depend on the frequency of a word relative to its friends and enemies rather than the number of friends and enemies.

In sum, Jared et al. found that words with high frequency friends are easy to pronounce and words with high frequency enemies are hard to pronounce. These results are most compatible with analogy and PDP models of word recognition. In an analogy model (e.g. Glushko, 1979) both neighborhood size and neighborhood frequency influence the synthesis process. Neighborhood size describes the number of items to be synthesized and neighborhood frequency describes their relative activations. Higher frequency items have stronger representations, which dominate in the synthesis of a pronunciation. In a PDP model (e.g. Seidenberg & McClelland, 1989), the associations which are formed within the network are influenced by the number of times a correspondence is encountered. Therefore, both neighborhood size and neighborhood frequency could influence the formation of attractors in the network. The assumption that allows the PDP models to account for neighborhood frequency effects is that during training, weights in the network associated with a particular word are changed more for an encounter with the word itself than for an encounter with an orthographic neighbor. "The factor that has the biggest effect on performance is the number of exposures to the word itself" (Jared et al., 1990, p. 709). So, high frequency words develop strong attractors, which will tend to interfere with the generation of pronunciations for low frequency words.

These results are least compatible with dual route GPC models, for which they present two potential problems. The first potential problem is that in Experiment 1, Jared et al. found neighborhood frequency effects among words that they claim were largely GPC regular. These findings would be difficult to accommodate within a dual route framework. However, examination of the stimuli from Experiment 1 reveals that the stimuli were not largely regular words; nearly 25% violate Venezky's (1970) GPC rules. The neighborhood frequency effects themselves present a more serious difficulty for dual route models. The assembled phonology mechanism in dual route models does not distinguish between type and token frequency. Repeated presentations of the same spelling-to-sound correspondence in different words and repeated presentations of a single word will cause an equivalent change in rule strength. This happens because the single letter contexts in both cases are the same. Consequently, there is no obvious way to produce neighborhood frequency effects in a GPC assembled phonology mechanism. Because of the limitations of the GPC hypothesis, and strong evidence for consistency, Jared (personal communication, November 11, 1995) concludes that neighborhood consistency is the only factor relevant to naming performance. According to Jared, GPC regularity effects are an artifact of failure to control for neighborhood consistency.

The evidence from the naming task reviewed above does not clearly indicate whether naming latencies are best predicted by GPC regularity, as dual route models suggest, or neighborhood consistency, as analogy and PDP models suggest. That is, we do not know whether "pint" is hard to pronounce due to on-line interference from "mint" , "lint" , and "hint" within a single assembled phonology mechanism or whether "pint" is hard to pronounce because retrieved and assembled phonology deliver conflicting pronunciations. Two critical questions can be asked to decide among these approaches. The first question is whether or not spelling-to-sound relationships among word bodies are predictable enough to do away with retrieved phonology. If they are, the second question is whether or not human subjects actually use them. Do readers adopt a single route or dual route solution to the problem of irregularities in English?

That is, do readers respond to irregularities by using larger units in the assembly process, applying that process to all words, or do they give up on assembly and rely on retrieval?

Word Body to Sound Relationships

When the orthographic neighborhoods of English are analyzed at the level of word bodies, as analogy and PDP models suggest, the relationship between spelling and sound is more predictable than when it is analyzed at the level of GPCs. This claim is supported by a statistical analysis of English words by Treiman, Mullenix, Bijeljac-Babic, and Richmond-Welty (1995), who computed various neighborhood statistics based on a sample of 1,329 Consonant-Vowel-Consonant (CVC) words. Treiman et al. first examined the consistency of the units that make up a CVC word when considered by themselves, without context. Consistency was computed in both a frequency independent fashion, which Treiman et al. refer to as a type measurement, and a frequency dependent fashion, which Treiman et al. refer to as a token measurement. The units for which calculations were done were the 3 parts of a CVC word. A CVC word such as "mint" consists of three units: the initial consonant, or C_1, "m" followed by the vowel, or V, "I", and the final consonant or consonant cluster, C_2 "nt". They found that the consistency of pronunciation of the C_1, the initial consonant, was relatively high (94% by types, 96% by tokens), the consistency of the V, the vowel units was relatively low (62% by types, 51% by tokens), and the consistency of C_2, the final consonant units, was fairly high (92% by type, 91% by tokens). This confirmed that vowel units are the source of many of the irregularities in English spelling-to-sound relationships. Vowel pronunciations were relatively more consistent with the addition of the C_2 as context, creating a VC_2 unit for which consistency was relatively higher than the vowel alone (80% by type, 77% by tokens). Vowel pronunciations were not more consistent with the addition of the C_1, creating a C_1V unit for which consistency was actually lower than for the vowel alone (55% by type, 52% by tokens). Similar results were found using all neighbors of words (both mono- and multi-syllabic ones). These findings confirmed the prediction that the spelling-to-sound relationships in English become more predictable when larger units are used. Furthermore, the unit must include letters following the vowel, as is the case with whole word and rime neighbors. Thus, it is possible that adopting larger units of a particular kind could make English more predictable.

Seidenberg et al. (1994) offer further support for the idea that spelling-to-sound relationships defined at levels other than GPCs are predictable enough to do away with retrieved phonology. They compared the word naming accuracy of the Plaut and McClelland (1993) version of their PDP model with the performance of the Coltheart et al. (1993) GPC rule simulation. Both models were tested on their accuracy at naming 2,897 words in the Seidenberg and McClelland (1989) corpus. Coltheart et al. reported that their model correctly pronounced 78% of the words in the corpus the 22% it missed were classified as exceptions by Coltheart. A dual route model, which Coltheart has not yet implemented, would rely on retrieved assembly to correctly pronounce these words. Plaut and McClelland (1993) reported that their PDP model correctly pronounced 99.7% of the words in the corpus -- this includes the exceptions. Their PDP model is sensitive to neighborhood consistency because it does not restrict context sensitivity to single grapheme neighbors. To the extent that the Seidenberg and McClelland (1989) word corpus represents a typical reader's core vocabulary, comparison of the performance of the PDP and GPC simulations in naming offers dramatic support for the idea that phonological recoding schemes other than GPCs are capable of pronouncing the majority of words in a reader's vocabulary.

GPC Regularity vs. Neighborhood Consistency

Treiman's findings and Seidenberg's comparisons support the idea that the problem of spelling-to-sound relationships in English can be solved by analyzing these relations at a level where they are more regular. The questions still remain whether (or not) this is the way that readers do solve this problem, and whether (or not) all readers adopt the same solution. Empirically, this question amounts to whether subjects' naming latencies are affected by GPC regularity or by neighborhood consistency, and whether these effects can be traced to the operation of one mechanism or two mechanisms.

Analyses of the properties of orthographic neighborhoods support the idea that units larger than graphemes are the basis of consistency effects. Treiman et al. (1995) compiled a list of stimuli which differed in GPC regularity, determined by Venezky's (1970) rules, and neighborhood consistency, determined by an onset/rime analysis of words in a subset of Kucera and Francis (1967). They found that having a consistent word body was a better predictor of naming latencies than following a GPC rule. Treiman argues that a number of sources of evidence, including speech errors, how phonemes are distributed in syllables, and people's ability to learn word division games that break syllables at various points, support the hypothesis that the unit of analysis in naming is larger than a correspondence between an individual grapheme and phoneme, instead corresponding to an onset and rime (see Treiman & Zukowski, 1988, for a review).

The conclusion that proponents of single route phonological recoding models would like to draw from this evidence is that the interaction between frequency and regularity is actually an interaction between frequency and consistency. The effects of GPC regularity on naming would then be an artifact of failure to control for neighborhood consistency (an idea proposed by D. Jared, personal communication, November 11, 1995). However, the case against dual route GPC models is inconclusive because it is based on evidence for sensitivity to consistency in naming rather than direct evidence against sensitivity to GPC regularity. Furthermore, it is possible that the system could be sensitive to both factors. According to Carr and Pollatsek (1985), the difficulty in deciding between models stems from the fact that the accounts are functionally equivalent -- they explain the same data in different ways. The GPC hypothesis is only one area in which the dual route and single route models differ. Hope for resolution of the debate comes from evaluating two other principles of the dual route model: independent processes and delayed phonology.

Testing the Dual Route Explanation

In a dual route model, frequency and regularity interact because for low frequency words the relatively slow assembled phonology routine competes with an independent retrieved phonology route. This account can be tested because dual route models predict that the effects of regularity on low frequency words should be alleviated if the assembly of phonology is interfered with more than the retrieval of phonology. That is, if it were possible to further delay the output of assembled phonology, it should be delivered late enough not to interfere with the retrieval of phonology of low frequency exception words. Paap and Noel (1991) tested this prediction by having subjects name words during the retention interval of a Sternberg (1966) memory task. The memory task served to create attentional interference of either a low magnitude, with a 1 item memory load, or a high magnitude, with a 5 item memory load. The imposition of a 5 item memory load speeded up the naming of low frequency exception words, exactly those words that dual route theories claim are named slowly due to interference, and slowed down the naming of all other types of words -- they exhibited what Bernstein and Carr (1996) call a release from competition (RFC) between the two routes from spelling to sound. Among the types of words that slowed with increased memory load; the largest slowdown was for low frequency regular words. This indicates a loss of redundancy gain (LoRG), low frequency words that were named quickly due to the simultaneous delivery of two redundant pronunciations are named slowly with increasing load because the assembled pronunciation was lost. Taken together these two effects provide strong support for the interpretation that concurrent task demands handicap the operation of the assembled phonology mechanism in a dual route model.

The dual task naming paradigm is interesting because it can be employed to answer the two questions which are central to the regularity vs. neighborhood consistency debate: (1) whether subjects' naming latencies are affected by GPC regularity or by neighborhood consistency, and (2) if these effects can be traced to the operation of one mechanism or two mechanisms. Existing research on the dual task naming paradigm has focused on the second possibility in that the effect has been used to differentiate between dual and single route explanations of the frequency

* regularity/consistency interaction in naming. Dual route models explain the effect as arising due to the release from competition between two functionally independent routes from spelling-to-sound and their explanation of the RFC and LoRG effects follows naturally and easily from this architectural account. However, single route phonological recoding models of naming might also be able to explain the RFC and LoRG effects.

Single Route Explanations of the RFC & LoRG Effects

Lukatela and Turvey (1993) devised a single route explanation of Paap and Noel's dual task naming results based on a series of 3 naming experiments using high and low frequency words and pseudohomophonic nonwords. Single route phonological recoding models predict that in Paap and Noel's dual task naming paradigm a word, such as "hope", and its pseudohomophone, "hoap", should be affected by load in the same way. This means that if one type of word is speeded up by load, its corresponding set of pseudohomophones should also be speeded because words and pronounceable nonwords are processed by the same mechanism. Dual route models predict that all pseudohomophones should display increased latencies with increasing load, even if their corresponding real words speeded up. This happens because pseudohomophones must be processed by the nonlexical system that is supposed to be slowed or inhibited by the load. Lukatela and Turvey found in 3 separate experiments that naming latencies for high frequency words and their pseudohomophones increased with load and naming latencies for low frequency words and their pseudohomophones decreased with load. They interpret this speedup as an indication that subjects usually verify responses in naming, an attention demanding process. Furthermore, this verification process slows the naming of low frequency words. An attention shortage caused by the memory load reduces the efficacy of this verification process so that subjects loosen their criterion with verification. This looser criterion benefits the naming of low frequency words and their pseudohomophones, which are generally verified slowly. The looser criterion does not benefit the naming of high frequency words and their pseudohomophones, which are generally verified quickly.

While this criterion shifting explanation may be a valid explanation of the results of Lukatela and Turvey's three experiments, there are two reasons it may not apply to Paap and Noel's results. First, the use of pseudohomophones, which essentially are misspelled words, could have caused subjects to engage in a spelling check in order to insure that they have accurately encoded the items -- this is verification. This strategy would not be required when subjects could be more certain of encoding accuracy, as when there are no pseudohomophones in the list. Therefore, criterion shifting as envisioned by Lukatela and Turvey would not necessarily be a strategy used in other dual task studies. The second problem is that Lukatela and Turvey do not report that they controlled for either the consistency or regularity of their naming targets. Inspection of their stimuli suggests that their naming targets are largely GPC regular. Their effect then differs from the RFC effect because the speedup is in the naming of low frequency regular words. In Paap and Noel (1991) and Bernstein and Carr (1996), it was the naming of low frequency exception words that speeded up with increasing load the naming of low frequency regular words significantly slowed with increasing load in both of these experiments. In conclusion, Lukatela and Turvey have demonstrated a speedup in naming latencies with load that may be explained by criterion shifting, but it is not the RFC effect.

Lexical analogy models of word pronunciation (Brooks, 1979; Glushko, 1979; 1981) could explain the RFC effect as arising due to cross-talk among lexical level candidates (However, it should be noted that this explanation is speculative.). In a lexical analogy model an irregular-inconsistent word is named slowly because its consistent neighbors compete with it during synthesis. This effect is stronger for low frequency words because frequency determines susceptibility to competition by strengthening representations. Strong representations (frequent and numerous ones) dominate during synthesis. A release from this competition for low frequency words could occur when the memory load in a dual task naming paradigm activates additional lexical candidates, which could dilute the interference that an irregular-inconsistent word suffers from its consistent neighbors. For this dilution to generate a speedup in naming latencies, the activation of competitors in an inconsistent neighborhood would need to be reduced more than the activation of the target itself. It remains to be proven whether or not an analogy model could actually generate this performance. These models would also need to explain the occurrence of the LoRG effect, which could happen if low frequency words were particularly susceptible to the activation of additional representations in memory.

A single route PDP model of word pronunciation (Plaut & McClelland, 1993; Seidenberg & McClelland, 1989; Van Orden et al., 1990) could also explain the RFC and LoRG effects as arising due to crosstalk but among distributed sub-lexical associations rather than among the lexical representations as in analogy models. These sub-lexical representations are encoding patterns, which Van Orden et al. (1989) describe as attractors. An attractor represents word specific covariance -- the tendency of certain orthographic and phonological patterns to co-occur. An irregular-inconsistent word is named slowly because inconsistent crosstalk among orthographically similar but phonologically distinct patterns pulls an encoding off its attractor. The effect of inconsistent crosstalk is stronger for low frequency words than high frequency words because the strength of an attractor increases when it is high in frequency. A release from this competition for low frequency words could occur when the memory load in a dual task naming paradigm activates additional attractors, which could dilute the interference which an irregular-inconsistent word suffers from other attractors. For low frequency consistent words, this would cause the LoRG effect. For this dilution to generate the speedup in naming latencies which marks the RFC effect, the activation of inconsistent attractors would need to be reduced more than the activation of the target's attractors.

Attention Demands or Cross-Talk?

Single route explanations of the RFC and LoRG effects by analogy and PDP models share the prediction that cross talk interferes with assembled phonology, which releases the system from interference for inconsistent words and causes additional interference for low frequency consistent words. This differs from the dual route account in which these effects are due to attentional handicapping of GPC rules. Bernstein and Carr compared the cross-talk account of the RFC and LoRG effects offered by analogy and PDP models with the attentional interference account of dual route models. This was accomplished by a between subjects comparison in the dual task naming paradigm of different kinds of memory loads that were argued to cause different kinds of crosstalk. A set of random shapes was argued to cause no crosstalk, just attentional interference, a set of digits and a set of nouns caused lexical crosstalk, and a set of CVC nonwords caused sub-lexical crosstalk. A calibration experiment established that the memory load caused attentional interference in all conditions.

Initial analysis of the data revealed a weak version of the appropriate pattern only in the digit and noun conditions -- weak support for the cross-talk explanation. Both slowdowns and speedups were much smaller than expected. Other attempts to replicate this effect have met with similar problems. Pexman and Lupker (in press) failed to replicate Paap and Noel's results, most notably the low frequency exception word speedup, even with methodological improvements aimed at increasing the magnitude of interference generated in their dual task procedure. In separate experiments, they tried a direct replication of Paap and Noel, increased emphasis on the memory aspect of the dual task paradigm, and enhanced demands in the high memory load condition; none of these manipulations resulted in a speedup in the naming of low frequency exception words. Strain and Patterson (personal communication, 1993) also report failures to replicate the RFC effect, and Herdman and Beckett (in press) also have had limited success. Herdman and Beckett did replicate the RFC and LoRG effects with memory loads of digits. Their failures came with memory loads of tones and random dot patterns. Based upon these failures to replicate, Pexman and Lupker concluded that Paap and Noel's original report was a case of Type I error. However, two alternative explanations of the failure to replicate have not been evaluated. One involves sampling error in the subjects; the other involves sampling error in the items.

Individual Differences In Susceptibility to Regularity Effects

Bernstein and Carr speculated that sampling error in the selection of subjects could underlie the failures to replicate Paap and Noel's results. [This runs counter to the modal opinion of the reading abilities of college students, whom the majority of researchers regard as a homogenous group.] This reasoning is based on the general belief that individual differences in susceptibility to regularity effects are due to variations in the age at which competence in spelling-to-sound knowledge is attained, and that schooling reduces individual differences. For instance, Waters, Seidenberg, and Bruck (1984) found that poor readers displayed greater sensitivity to regularity than good readers in both latencies and error rates in three reading tasks: naming, lexical decision, and semantic plausibility judgments for words finishing sentences. These differences were present in young children (2nd and 3rd grade), but not in older children (5th grade), for whom sensitivity to regularity was similar to that of college students. They interpreted these results as supporting the dual route model, in which there are two spelling-to-sound mechanisms, and that poor readers have impaired phonological recoding abilities. The implication of finding that 5th graders and college students display similar susceptibility to regularity is that once the system achieves its mature state all readers should perform similarly.

Contrary to popular belief and to the Waters et al. (1984) results, Morrison (1995) found in a longitudinal study that schooling did not reduce individual differences in the reading ability of grade school students. Evidence that individual differences continue even into college comes from studies of the relationship between print exposure and reading performance by Stanovich and West (1989), who found that individual differences in print exposure were correlated with the orthographic and phonological processing performance of college level readers. In Experiment 2, Stanovich and West classified their subjects as good or poor readers, based on their performance on the Woodcock Reading Mastery Tests. Good readers were significantly less susceptible than poor readers to regularity effects in naming, both in latencies and in error rates. Good readers were also less susceptible to regularity effects in spelling, which Stanovich and West used as a measure of orthographic processing skills.

Stanovich and West examined the relationship between reading experience and individual differences in reading performance by assessing print exposure with two measures. The first was a survey of reading and television habits. Because subjects tend to respond to surveys about reading with socially desirable answers, Stanovich and West also included an author and magazine recognition test as measures of experience free of this bias. Print exposure, as measured by performance on the author recognition test, was a good predictor of orthographic and phonological processing skills both in correlations and in a regression analysis. Other measures of print exposure also worked but not so well as the author recognition test scores.

Individual Differences & The RFC and LoRG Effects

Bernstein and Carr (1996) speculated that individual differences in susceptibility to regularity effects could be responsible for weak replications of Paap and Noel's results. That is, in order for a memory load to influence the magnitude of regularity effects by handicapping the assembled phonology mechanism, subjects must first display regularity/consistency effects in the dual task naming paradigm. Furthermore, the magnitude of these regularity/consistency effects must be relatively large. Assuming that the frequency by regularity/consistency interaction reflects normal or average performance in naming tasks, subjects who do not display the interaction may have somehow changed their naming performance. Exploratory data analysis revealed that the signature of susceptibility to regularity effects of the type predicted by the dual route architecture -- the appearance of the frequency by regularity/consistency interaction -- exhibited large amounts of within groups variation. In fact, only about half the subjects showed such an interaction pattern in their word naming latencies under low memory load, which is the control or "normal reading" condition in dual task naming studies. To explore the possibility that the appearance of the frequency by regularity interaction was related to performance under increasing memory load, a post-hoc division of subjects into groups was performed.

One group of subjects displayed the frequency by regularity interaction pattern in their mean naming latencies for the one item memory load condition of the dual task naming paradigm and the other group did not. Subjects were identified as displaying the frequency by regularity/consistency interaction if their mean naming latencies exhibited four characteristics: (1) High frequency words were named faster than low frequency words. (2) For low frequency words, regular words were named faster than exception words. (3) The difference between exception and regular word naming times for low frequency words was larger than the analogous difference observed for high frequency words. (4) The difference between exception and regular word naming times for low frequency words was larger than the median value. If subjects' mean naming latencies did not display these four characteristics, they were classified as not displaying the frequency by regularity interaction.

Classification of subjects by this method had a significant impact such that subjects who displayed the frequency by regularity interaction also displayed the RFC effect. Naming latencies for low frequency exception words speeded up with increasing memory load, whereas naming latencies for other types of words slowed down. Subjects who did not display the frequency by regularity interaction did not display the RFC effect. Naming latencies for all words slowed with increasing memory load. Neither group of subjects displayed the LoRG effect. Bernstein and Carr speculated that these might be caused by variation in system architectures. However, the differences in performance among groups of subjects could also be an artifact of the selection criteria. This possibility is explored in the second chapter.

CHAPTER 2

A Critical Analysis of Post-Hoc Subject Classification

Regression to the Mean

The classification method used by Bernstein and Carr was the post-hoc division of subjects into two groups or blocks based upon the appearance and magnitude of the frequency by regularity interaction in the one-item load condition of the dual task memory experiment. One criticism of this procedure is that the selection of subjects who display the frequency by regularity interaction could be equivalent to the selection of subjects who exhibit unusually long naming latencies for low frequency exception (LF-E) words. The resulting speedup in LF-E word naming latencies with increasing memory load could then be due at least in part to regression to the mean. Bernstein and Carr (1996)'s defense against this criticism was that classification of subjects by a median split of LF-E word naming latencies did not produce either the RFC or LoRG effects. However, this defense was based on an operational definition of a long LF-E word naming latency as one which falls in the upper half of the distribution of LF-E word subject means. This definition failed to consider that the base rate for naming latencies differs among subjects. Base rate naming latencies for LF-E words are typically established using LF-R word naming latencies. A long naming latency for a LF-E word would then be operationally defined as one for which the difference between latencies for LF-E and LF-R words falls in the upper half of its subject distribution, which is one of the selection criteria used by Bernstein and Carr. Therefore, if the combined criteria used by Bernstein and Carr end up selecting essentially the same subjects as those selected only on the magnitude of the difference between LFE and LFR words, then their RFC effect could still represent regression to the mean.

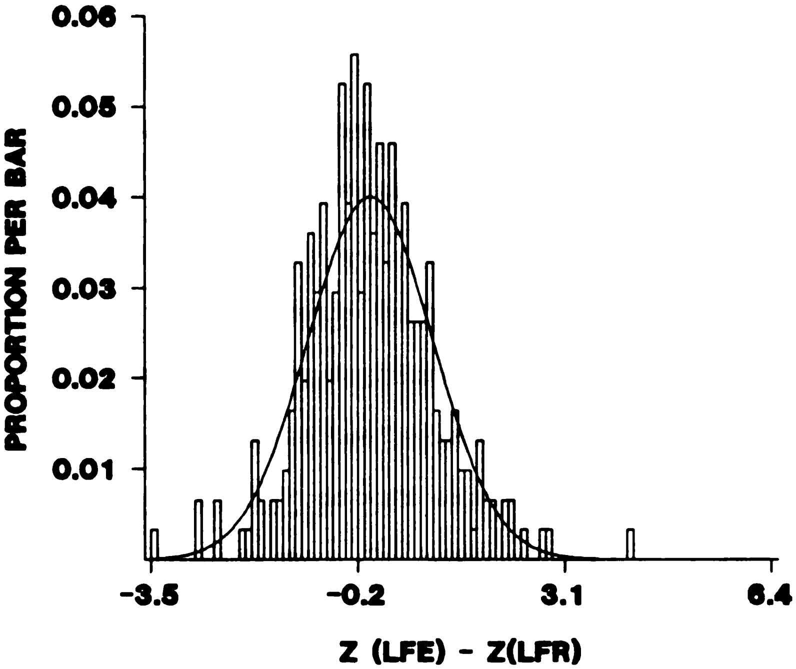
To see if this was the case, a correlational analysis was performed on the data from Experiments 2 and 3 from Bernstein and Carr (1996) to determine whether the use of both selection criteria, the appearance of the interaction and the magnitude of the LFE-LFR difference, differed significantly from considering the magnitude alone. A Pearson chi-square test for independence indicated that the difference between the two sample distributions of subjects was significant, $\chi^2 = 70.91$, $p < .001$, df=1. However, it was true that selecting subjects who met both criteria included many subjects with relatively large LF regularity effects, 75% to be exact. Consequently, the argument that selecting subjects by both criteria could not include some regression to the mean is not very strong.

Violations of Independence & Problems With Median Splits

A further problem with the classification procedure is that it violates independence. Subjects were divided into groups or blocks based upon a measure derived from subjects scores on a dependent variable. According to Keppel (1991), this practice can lead to an increase in type I error rate. Maxwell and Delaney (1993) describe how one method of dividing subjects into blocks, performing a median split, can lead to this kind of increase in Type I error rate. They argue that median splits can create a difference where none exists. A single sampling distribution is divided in half -- creating upper and lower local central values typical of bimodal distributions where there is actually only a single central value. Because the ANOVA is blind to the source of this introduced variance, it can cause spuriously significant F values. Therefore, applying a median split to a normal distribution will increase the type I error rate.

Two things need to be true for this problem to apply to the classification scheme used by Bernstein and Carr. First, classifying subjects by the pattern of thier naming latencies would need to be equivalent to performing a median split on a single dependent variable. Second, the distribution of scores for this dependent variable would need to be unimodal rather than bimodal. The shape of the distribution is important because bimodality would support the hypothesis that categorical differences among subjects exist. That is, the similarity between the pattern classification and median split would not be a problem if the distribution of low frequency regularity effects under a memory load of 1 item were bimodal. This would indicate a categorical difference among subjects in susceptibility to regularity effects. One peak of the frequency distribution would be at zero, representing those subjects who are not affected by regularity and one peak would be above zero, representing those subjects who are affected by regularity. Although Bernstein and Carr did not evaluate this hypothesis, it could be evaluated with a sufficient number of low frequency regularity scores. Combining data from the Bernstein & Carr (1996) Experiments 2 and 3 yields 275 scores. Because the mean and standard deviation of naming times differed among the lists of stimuli used in these experiments, difference scores were standardized before being combined.

Figure 2: Distribution of LF Regularity Effects



The frequency distribution in Figure 2 does not display any tendency towards bimodality. This result is not evidence against the hypothesis that subjects display individual differences in susceptibility to regularity effects, but it does suggest that these differences might not be categorical (qualitative) -- the differences are purely continuous (quantitative).

A continuous measure of the magnitude of the F*R Interaction

The previous section suggests that a categorical division of subjects who do and do not display the F*R interaction is not justified because there was no tendency towards bimodality in the distribution of low frequency regularity effects. Therefore, an ANOVA is unjustified in which apparent categorical individual differences in the appearance and magnitude of the frequency by regularity interaction are used as a blocking factor to predict the three way interaction of A (frequency) x B (regularity/consistency) x C (memory load), with subjects as the random factor. An alternative way to evaluate the possibility that individual differences in the magnitude of the frequency by regularity interaction predict the interaction of frequency, consistency, and memory load is to base the inferential analysis of naming latencies on a regression model rather than a blocking model.

There are two candidates for the individual differences variable to be used in the regression. One candidate is the magnitude of the low frequency regularity effect. However this measure does not completely capture the notion of an interaction, which the categorization criteria do. An alternative is the magnitude of the low frequency regularity effect minus the magnitude of the high frequency regularity effect. This continuous measure reflects the magnitude of the frequency by regularity interaction. This additional factor, hereafter called FBYR, varies between subjects. The factors in the general linear model analysis were A (Frequency) x B (Regularity) x C (Memory load) = D(FBYR), with subjects as the random factor. The model underlying this analysis is identical to the analysis of covariance, but the purpose of the additional factor was not to remove the effects of a confound that mask an effect of interest, as in the ANCOVA, but to determine the nature of the effects of the additional factor.

The inclusion of the individual differences factor in the general linear model analysis allows the direct assessment of the possibility that the magnitude of the frequency by regularity interaction under a 1 item memory load is a significant predictor of the magnitude of this interaction under a 5 item load, which will be indicated by the significance of the four factor interaction of A (frequency), B (regularity), C (load), and FBYR. The advantage of the regression solution over blocking is that no selection or division of subjects is necessary to assess the effects of individual differences. Therefore, neither the introduction of artifactual variance nor regression to the mean is a potential explanation of any effects of individual differences that might be found.

For the purpose of comparing the power of the regression and blocking solutions, both analyses were applied to the data from Bernstein and Carr (1996) Experiment 2. In the regression analysis, the continuous measure of the magnitude of the frequency by regularity interaction was based on the 1 item memory load naming latencies. In the blocking analysis, subjects were divided into groups according to the criteria explained previously. The blocking solution revealed a significant interaction of classification, frequency, consistency, and load, $F(3,159) = 15.552$, $p < 001$. The effect size of this interaction was relatively small, $eta^2 = 0.2269$. In the regression solution, the four factor interaction of the covariate with frequency, regularity/consistency, and memory load was highly significant, $F(1,175) = 155.41$, $p < .0001$, $MSe = 1806$, indicating that the magnitude of the frequency by regularity interaction under a 1 item load was a significant predictor of how it was affected by an increase in load. The effect size for this interaction was large, $eta^2 = 0.4746$. Therefore, the regression solution accounted for roughly twice the variance accounted for by the blocking solution.

The regression solution also increased the overall power for detecting the interaction of frequency, regularity/consistency, and load. The value of this interaction in the blocking design was $F(1,159) = 0.635$, $p > .05$ with an eta^2 of 0.0039. In the regression analysis, the overall interaction of frequency, regularity/consistency, and memory load using type III sums of squares was significant $F(1,175) = 20.16$, $p < .0001$, $MSe = 1806$, indicating that when the error variance due to individual differences in the magnitude of the F*R/C interaction was removed, the RFC effect was present. The effect size for this interaction was also increased to $eta^2 = 0.321$.

Summary of Critical Analysis

In this critical analysis, the F*R/C pattern and median split classification schemes were similar in that selecting subjects who displayed the F*R/C interaction included many of the subjects who had relatively large LF regularity effects. There was also a failure to find a bimodal distribution of low frequency regularity effect scores, which suggests that the individual differences in susceptibility to regularity effects may be continuous rather than categorical in nature. Based on both of these results, one cannot dismiss the possibility that the speedup in dual task naming latencies for LF-E words with increasing memory load which Bernstein and Carr (1996) report using their classification scheme included some regression to the mean.

A mixed model regression is a method for both isolating and assessing the effects of individual differences in susceptibility to regularity effects. This approach is not subject to the regression to the mean criticism, since no selection occurs. The results of the general linear model analysis roughly double the proportion of variance accounted for by the interaction of individual differences with frequency, regularity/consistency, and memory load. This supports the hypothesis that the magnitude of the frequency by regularity interaction under a 1 item memory load is a significant predictor of the frequency by regularity by load interaction. However, the results of this analysis should still be interpreted with a certain amount of caution, since the individual differences variable was not independent of the other factors. An attempt at alternative analyses which do not have this limitation will be made in the experiments which follow.

Introduction to the Experiments

During reading, competition among candidate pronunciations causes the ambiguity present in spelling-to-sound relationships to influence response latencies in the naming of printed monosyllabic words. Paap and Noel (1991) found that the imposition of a memory load during naming could reduce the impact of this competition by speeding naming latencies for low frequency GPC exception words while slowing naming latencies for other types of words. This result has been interpreted as supporting dual route GPC models of pronunciation (Bernstein & Carr, in press; Herdman et al., in press; Paap & Noel, 1991; Paap et al., 1992). However, difficulty in replicating this result has been interpreted as a failure to confirm the predictions of dual route GPC models (Pexman & Lupker, 1995; Strain & Patterson, personal communication). In a post-hoc analysis, Bernstein and Carr (1996) argued that failures to replicate were due to failures to account for individual differences in susceptibility to GPC regularity effects. This finding is in need of systematic investigation to establish the role of individual differences in predicting the effect of memory load on naming latencies. A further issue which needs to be resolved if the effect proves to be reliable is whether a regularity or consistency based explanation of the effect is correct.

Individual Differences and the RFC Effect

Bernstein and Carr (1996) argued that failures to replicate Paap and Noel's (1991) results were caused by individual differences in susceptibility to regularity effects, which manifested themselves as variation in the appearance or magnitude of the frequency by regularity interaction. An alternative explanation of individual differences in the appearance or magnitude of the frequency by regularity interaction in dual task naming is that they could have been a result of the way in which performance changes in a dual task situation. When memory span and naming are executed concurrently, performance will be different than it is when the tasks are executed separately. Performance will change in both intended and unintended ways. The intended changes are easy to detect because experiments were designed to reveal them. In Paap and Noel's dual task naming paradigm, the intended changes are the differences in naming latencies as memory load is increased from 1 item to 5 items. The unintended changes are harder to detect because they can involve the violation of untested assumptions. For instance, researchers have assumed that a 1 item memory load would not cause sufficient interference to qualitatively change naming performance. It is possible that the large amount of between-subjects variation in 1 item memory load naming performance that Bernstein and Carr identified was an unintended change relative to single task naming performance. Single task performance would then be much more homogenous than dual task performance. It is also possible that this amount of between-subjects variation exists in single-task naming latencies.

A direct comparison of single and dual task naming was performed in Experiment 1 to help determine whether individual differences in the appearance of the frequency by regularity interaction in dual task naming performance generalize to single-task naming. In the event that they do, classification of subjects in dual task naming experiments could be based on independently collected single task naming data. This would avoid the potential problems with classification based on performance in the low-load condition of the dual task paradigm. A related issue addressed in Experiment 1 was whether a continuous or categorical measurement of the frequency by regularity interaction would be more reliable. The purpose of this comparison was to establish which kind of measurement should be used to capture individual differences in the frequency by regularity interaction.

Information on print exposure was also collected in Experiment 1 in the form of scores on an author recognition test, a magazine recognition test, and a reading and media habits questionnaire. Together this information can help to identify the role of experience in the appearance of individual differences in the magnitude of the frequency by regularity interaction. Furthermore, if these measures are highly correlated with individual differences, they can be used to select subjects. This would have the advantage of being a measurement independent of naming performance.

Regularity or Consistency

The second issue that was explored in the experiments is single vs. dual route explanations of the RFC effect. In existing dual task naming experiments investigating the RFC effect, GPC regularity and neighborhood consistency have been confounded in the lists of naming stimuli. Consequently, both single and dual route explanations of the RFC effect are possible. In Experiments 2 and 3 of this dissertation, these stimulus properties were partially unconfounded so that each explanation could be independently tested.[2] According to dual route GPC models the RFC effect is the result of lessening the interference that assembled phonology causes in the naming of a low frequency GPC exception word. This possibility was tested in Experiment 2. According to analogy and PDP models, the RFC effect is the result of diluting the interference caused by a higher frequency inconsistent neighbor of a low frequency word. This possibility was tested in Experiment 3.

[2] These properties overlap to the extent that complete unconfounding was impossible.

CHAPTER 3

Experiment 1

The purpose of Experiment 1 was three-fold. a) To determine the most reliable method of identifying individual differences in the appearance or magnitude of the frequency by regularity interaction; b) To determine whether these differences generalize to single task naming or are unique to dual task naming; and c) To determine the relationship between reading experience and the appearance or magnitude of the frequency by regularity interaction. The answers to these questions helped identify the least error prone method for removing the effects of individual differences in dual task naming. To answer these questions single and dual task naming performance was compared within the same group of subjects using a list of high and low frequency GPC regular and exception words. The list of naming stimuli was as long as possible to allow the validity of subject classification methods to be assessed in a split-half analysis. The dual task naming consisted only of the one item memory load condition. The five item memory load condition was not included because it was not directly relevant to the identification of differences in susceptibility to regularity effects. Comparisons between the single and dual task naming paradigm allowed the generalization of individual differences to be assessed. Finally, data on reading experience were also collected for each subject, using the questionnaire and survey from Stanovich and West (1989). This data allowed the exploration of the relationship between reading experience and the appearance or magnitude of the frequency by regularity interaction in naming.

Method

Subjects

Twenty four undergraduate psychology students were recruited from the subject pool at Michigan State University and received class credit as compensation for their participation. Participation was restricted to native speakers of English with normal or corrected-to-normal vision. The number of subjects necessary was determined by power analysis (Cohen, 1977). The effect size was estimated using the frequency * regularity * subject classification from Bernstein and Carr (1996) Experiment 3, where f = .5 (f is Cohen's measure of treatment magnitude). For the design of this experiment with α = .05 and u = 3, Cohen's (1977) tables indicate that 24 subjects results in power in excess of .95, which exceeds Cohen's recommendation of power of at least .80 in psychology experiments.

Apparatus

All stimuli were presented and all responses except naming errors (which were scored by the experimenter) were collected using an IBM compatible computer with a Magnavox EGA monitor. A Gerbrands model G-1341 voice activated relay was attached to game port as an analog switch closure. A microphone placed on a stand was attached to the voice activated relay and a separate microphone was attached to a tape recorder used to record the sessions. The presentation of stimuli and the recording of all reaction times was controlled by a program written in Microsoft Quick Basic. Memory responses were made on a Microsoft-compatible two button serial mouse with the track ball removed. The left button was marked "yes" and the right button was marked "no".

Materials

Memory load stimuli consisted of the digits stimuli from Paap and Noel (1991). Naming stimuli consisted of 208 items, 52 in each of 4 categories: high and low frequency GPC regular and exception words. The repetition of word bodies was avoided, e.g. mint and pint were not both used. The list itself was compiled from 3 sources. A list appears in Appendix B, consisting of 112 items; 96 items from Taraban and McClelland (1987) and an additional 16 items from Paap and Noel (1991). The repetition of word bodies in the combined list was avoided. Finally, a list of 96 words with non-productive (unique) word bodies, which appears in Appendix C, was used.

Because no other words share these words bodies, they allow the list of naming stimuli to be made considerably longer without the confound of repetition.

The reading experience and media habits questionnaire, the author recognition test, and the magazine recognition test, which appear in Appendix A, were taken from Stanovich and West (1989). The list of 50 non-famous authors, which serve as the foils in the author recognition test, was taken from the editorial board and a handful of authors from the Journal of Experimental Psychology: Learning, Memory, and Cognition, 21(1). The names of editors from Michigan State University were not used in order to avoid a confound of familiarity.

Procedure

For each subject, the list of 208 naming stimuli was randomized within the categories high and low frequency regular and exception before being divided into two lists of 104. One list of 104 items was combined with the memory load stimuli for the dual task naming condition. The other list of 104 items was used for the single task naming condition. No naming stimuli were seen by the same subject twice. This method of dividing the stimuli was intended to avoid any confound due to the selection of particular items for the single and dual task naming conditions. In the dual task naming condition, naming stimuli were paired with the memory load items in a different random drawing for each subject to avoid any particular naming item always being paired with a particular memory stimulus.

Subjects were told that they were participating in a series of three tasks intended to explore whether memory and reading interfere with each other when performed concurrently. One condition was dual task naming under the guise of a memory span experiment. The other condition is single task naming. The order of these first two conditions was counterbalanced between subjects. The final condition for all subjects was a brief survey. Subjects were individually tested in a small room. They sat in front of the PC at a distance that varied between approximately 40 and 80 cm. Under these conditions, the visual angle subtended by a four letter

18 point stimulus varies from 2 degrees 8 minutes to 1 degree four minutes, and the visual angle subtended by a four letter 24 point stimulus varies from 2 degrees 12 minutes to 1 degree 55 minutes.

The dual task portion of the experiment consisted only of 1 item memory load trials. A dual task naming trial began with the visual presentation of a fixation point for a duration of 1000 ms that was in turn followed by the presentation of one randomly selected item for the memory load task. The study item remained on the screen for 400 ms, after which the naming stimulus appeared following a delay which varied between 1 and 2 seconds after the offset of the memory set, during which time a blank screen was displayed. The randomness of the delay was intended to eliminate any planned switching of attention to the naming task. The naming stimulus remained on the screen until the subject responded. After the response a blank screen was displayed until 4 seconds elapsed from the offset of the memory set. This blank screen was followed by the presentation of the memory probe item that remained until the subject responded by either pressing the left mouse button, marked "yes", if the item had appeared in the study set or by pressing the right mouse button, marked "no", if the item had not appeared in the study set. Errors in memory responses resulted in the sounding of a warning tone, followed by a one second presentation of a screen containing the message "error -- please be more careful." An inter-trial interval of 3 seconds elapsed between the memory response, or the offset of the error message, and the beginning of the next trial. Subjects completed 4 practice trials that contained memory load stimuli identical to those used in the experimental trials followed by the 104 experimental trials, which began with 2 filler items. The program paused every 26 trials so that subjects could take a break.

A single task naming trial began with the presentation of a fixation point, "+" centered on the screen for 500 ms, immediately followed by the naming stimulus, which remained on the screen until the subject named the word. This was followed by a 3 second inter trial interval.

Subjects completed 4 practice trials followed by the 104 experimental trials, which began with 2 filler items. The program paused every 26 trials so that subjects could take a break.[3]

Following completion of the single and dual task naming trials, subjects were asked to fill out the reading and media habits questionnaire and to take the magazine and author recognition tests. An experimental session lasted approximately 45 minutes.

Results

Naming latencies were discarded from trials on which there were pronunciation errors (3.4%) or voice key triggering errors due to environmental noises (1.4%). Latencies 125 ms or shorter were also regarded by the experimenter as voice key triggering errors (< 1%). This figure of 125 ms is comparable to cutoffs used by other investigators (125 ms -- Bernstein & Carr, in press; Taraban and McClelland, 1987; 120 ms -- Paap and Noel, 1991). Naming latencies longer than 2,000 ms were scored as null responses (1%). Errors in naming responses were determined by the experimenter who observed the session, and uncertainties were resolved by listening to the audio tape. In the dual task condition, naming latencies were discarded from trials containing memory errors (1.6%).

To achieve a long list of stimuli, words with productive and non-productive bodies were combined for this experiment, with the assumption that both types of words are affected similarly by memory load. This assumption was evaluated by analyzing naming latencies for correct pronunciations only in a 2 (frequency) x 2 (regularity) x 2 (memory load) x 2 (productive/non-productive word body) repeated measures ANOVA. There was neither an interaction of word body type and memory load nor any higher order interactions involving these two factors. This indication that memory load has similar effects on words with productive and non-productive

---

3 Unlike the dual task naming trials, the interval between the fixation point and the naming response was fixed. While a variable onset of the naming stimulus would have more closely matched the dual task naming procedure, it would also have reduced the reliability of the frequency by regularity interaction.

bodies justified the collapsing of words with productive and non productive bodies into larger categories.

Overall Analysis of Naming Latencies & Error Rates

An analysis of variance was performed to test the hypothesis that the frequency by regularity interaction in single task naming was changed by dual task naming. Naming latencies were averaged across items and analyzed with subjects as the random factor in an ANOVA with frequency(2), regularity (2), and task (2) as within groups factors. Naming latencies were also averaged across subjects and analyzed with items as the random factor in an ANOVA with task (2) as a within groups factor and frequency (2), as well as regularity (2) as between groups factors. Latencies, which are graphed in Figure 3, displayed the standard frequency * regularity interaction pattern in both single and dual task naming. Subjects named high frequency words (631 ms) significantly faster than low frequency words (680 ms), $F(1,23)=33.565$, $p < 0.001$, $MSe = 3434$ by subjects, $F(1,204)=27.602$, $p < .001$, $MSe=8025$ by items. Regular words were named (645 ms) significantly faster than exception words (665 ms), $F(1,23) = 18.093$, $p < 0.001$, $MSe = 1102$ by subjects; $F(1,204) = 4.700$, $p < .05$, $MSe=8025$ by items. There was a significant interaction of frequency and regularity by subjects, $F(1,23)=23.205$, $p < .001$, $MSe=1383$ but not by items $F(1,204) = 1.277$ $p > .05$, $MSe=8025$, such that low frequency exception words were named 46 ms slower than low frequency regular words whereas high frequency exception words were named 5 ms faster than high frequency regular words.

The impact of the dual task paradigm on naming latencies was to cause an overall slowdown and a decrease in the size of the frequency effect; regularity effects were uninfluenced by memory load. Subjects named words significantly more slowly in the dual task condition (731 ms) than in the single task condition (579 ms), $F(1,23) = 57.809$, $p < .001$, $MSe=19326$ by subjects, $F(1,204) = 817.755$, $p < .001$, $MSe = 3156$ by items. There was a significant interaction of task and frequency by subjects, $F(1,23) = 4.751$, $p < .05$ but not by items, $F(1,204) = 0.407$, $p >$

.05, MSe=3156, such that high frequency words were named 57 ms faster than low frequency words in single task naming while the corresponding difference in dual task naming was 41 ms.

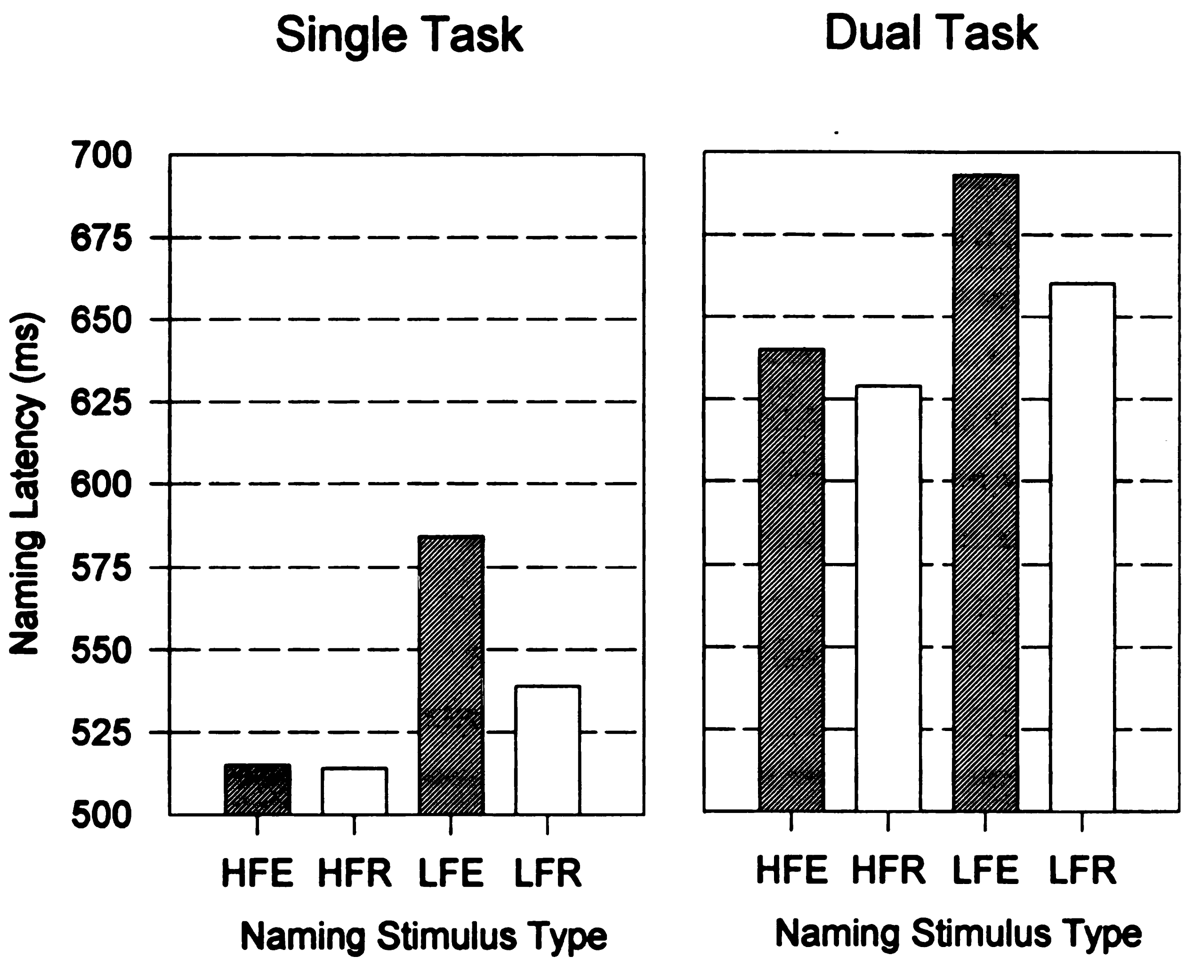


Figure 3: Frequency * Regularity Interaction For Single vs. Dual Task Naming

	High Freq. Exception	High Freq. Regular	Low Freq. Exception	Low Freq. Regular
Single Task Naming	0.54	0.29	2.25	0.63
Dual Task Naming	0.29	0.38	2.04	0.25

Table 1: Proportion of errors in naming.

Naming error rates, which are reported in Table 1, were analyzed in the same fashion as naming latencies. Error rates were significantly greater for naming low frequency (1.292) vs. high frequency (0.375) words, $F(1,23)=37.608$, $p < .001$, MSe=1.072. Error rates were significantly greater for exception (1.282) vs. regular (0.386) words, $F(1,23) = 55.446$, $p < .001$, MSe=0.695. There was a significant interaction of frequency and regularity, $F(1,23) = 30.606$, $p < .001$, MSe=1.035, such that the effect of regularity on error rates was greater for low frequency (1.833) than high frequency (0.083) words. Task had no significant effect on naming error rates.

Subject Classification Analyses

Internal Consistency of Continuous Measures

Another purpose of this experiment was to find a reliable measure of the appearance or magnitude of the frequency by regularity interaction in naming. For each subject, 5 continuous measures were calculated that could potentially reflect differences in susceptibility to regularity effects: (1) Mean low frequency exception word naming latency. (2) A continuous measure of the overall regularity effect: (LFE + HFE) - (LFR + HFR). (3) A continuous measure of low frequency regularity effects: (LFE - LFR). (4) A continuous measure of high frequency regularity effects: (HFE - HFR). (5) A continuous measure of the frequency by regularity interaction: (LFE - LFR) - (HFE - HFR). The reliability of each variable was assessed by performing a split-half analysis of the latencies from the single task naming condition. Because the list of naming stimuli was twice as long as the typical list used in dual task naming studies the split-half analysis should not underestimate the stability of the variables. Correlations for this analysis appear in Table 2 and the data on which this analysis was performed appear in Appendix E. Only single task naming latencies were used to avoid the possibility that dual task or memory relevant factors would influence naming performance. The correlations for two factors that do not reflect differences in susceptibility to regularity effects also appear in Table 2. These factors are (1) mean latency and (2) the frequency effect: (LFE + LFR) - (HFE + HFR). Including these factors should help establish the validity of the split-half analysis because these variables should be relatively stable.

	LFE Latency	LF Regularity (LFE - LFR)	HF Regularity (HFE - HFR)	Regularity (LFE + HFE) - (LFR + HFR)	F*R (LFE-LFR) - (HFE - HFR)	Mean Latency	Frequency (LFE + LFR) - (HFE + HFR)
r	.8925	.4018	.6304	.5156	.3675	.9563	.5606
p	.0001	.0516	.0001	.0099	.0773	.0001	.0044

Table 2: Results of the split-half reliability analysis for latency measures.

The check on the split-half analysis, the mean latency and frequency effect variables, were both significantly correlated between the two halves of the list. Almost all the continuous indicators of susceptibility to regularity effects were significantly correlated between the two halves of the list. The only exception was the F*R interaction variable, which approached significance.

Internal Consistency of Categorical Measures

For each subject, four categorical measures were calculated that could potentially reflect differences in susceptibility to regularity effects: (1) Whether the frequency by regularity pattern was present or absent. This measure was scored as described earlier. (2) Whether low frequency regularity effects were above or below the median value. Subjects who scored above the median value were given a score of 1, subjects who scored below the median value were given a score of 2. (3) A combined categorical measure of both the appearance of the F*R interaction from #1 and the magnitude of low frequency regularity effects from #2. (4) Whether the continuous measure of the F*R interaction from Table 2 was above or below the median value.

A Pearson chi-square test for goodness of fit was calculated for these variables. Values are reported for those cases in which the expected frequencies were high enough for the chi-square test to be valid. Fisher's exact test was also calculated because expected frequencies were too small to use the chi-square approximation for two of the variables in Table 3. Fisher's exact test gives "the exact probability for a sample showing as much or more evidence for association than that obtained, given only the operation of chance." (Hays, 1994, p. 863). Values for these statistics appear in Table 3 and the data on which this analysis was performed appear in Appendix E.

	F*R Present / Absent	LF Regularity Above/Below Median	F*R Present/Absent & LF Regularity Above/Below Median	Continuous Measure of F*R Interaction Above/Below Median
% Agreement	62.5	50.0	37.5	50.0
Chance	50.0	50.0	25.0	50.0
χ^2	x	.000	xx	.000
p		1.000		1.000
Fisher's exact test	.647	1.000	.434	1.000

(x = 50% of the cells have expected counts of less than 5.)
(xx = 95% of the cells have expected counts of less than 5.)

Table 3: Results of the split-half reliability analysis for categorical measures.

The most reliable of the categorical measures was the presence/absence of the F*R interaction, which was in agreement in 62.5% of cases (chance agreement was 50%). However, according to Fisher's exact test, this amount of agreement is highly likely (p=.65) under the null hypothesis that classification based on the two halves of the list are not associated. The combined categorical measure was in agreement more than chance, but not significantly so. Neither of the median split variables was more reliable than chance.

Independent Predictors of the F*R Interaction

Another goal of the experiment was to find a variable to serve as an independent predictor of the appearance and magnitude of the frequency by regularity interaction. A multiple correlation analysis was performed in order to determine the best independent predictor of susceptibility to regularity effects in dual task naming. Measures of print exposure were first compared with four different measures of dual task naming performance. Four measures of print exposure were calculated for each subject: (1) ART: author recognition test (2) MRT: magazine recognition test (3) READING: self reported reading experience (4) TV: self reported television viewing. These measures were scored as indicated by Stanovich and West (1989). The measures of dual task naming performance that were used were based on the continuous measures that had proven to be reliable (or nearly so in the case of the continuous measure of the F*R interaction) in the split-half analysis. These four measures were scored as indicated earlier. Table 4 displays the Pearson correlation matrix of these 4 measures of experience with dual task naming performance.

	Experience Measures			
Dual Task Measures	Reading	TV	ART	MRT
F*R	-0.14	0.07	-0.10	-0.04
Frequency	-0.24	0.27	-0.54 **	-0.27
Regularity	-0.17	0.49 *	-0.26	-0.32
LF Regularity	-0.17	0.28	-0.19	-0.19

Table 4: Pearson correlation matrix for experience measures and dual task naming performance. (* $p < .05$; ** $p < .01$, *** Bonferroni adjusted $p < .05$)

None of the experience measures was significantly correlated with the continuous measure of the F*R interaction. Experience was correlated with other measures of dual task naming performance. Author recognition test scores were significantly correlated with frequency effects such that as recognition test scores became higher, frequency effects became smaller. Self reported television viewing was significantly correlated with overall regularity effects such that as TV viewing increased (and reading presumably decreased), regularity effects became larger.

A separate multiple correlation analysis examined single task naming measures as predictors of dual task naming performance. There were eight measures used: (1) F*R: the size of the frequency by regularity interaction. (2) Freq: the size of the frequency effect. (3) Reg: the size of the regularity effect. (4) LF-Reg: the size of the low frequency regularity effect. (5) HF-E: high frequency exception word naming latencies. (6) HF-R high frequency regular word naming latencies. (7) LF-E: low frequency exception word naming latencies. (8) LF-R low frequency regular word naming latencies. Table 5 displays the Pearson correlation matrix.

	Single Task Measures							
Dual Task Measures	F*R	Freq	Reg	LF-Reg	HF-E	HF-R	LF-E	LF-R
F*R	-0.04	0.39	0.15	0.05	-0.19	-0.27	-0.01	-0.06
Frequency	0.30	0.69 ***	0.49 **	0.42 *	0.24	0.17	0.49	0.36
Regularity	-0.02	0.49 *	0.10	0.04	-0.06	-0.11	0.12	0.14
LF Regularity	-0.04	0.49 *	0.14	0.05	-0.15	-0.23	0.05	0.03

Table 5: Pearson correlation matrix for single and dual task naming performance. (* $p < .05$; ** $p < .01$, *** Bonferroni adjusted $p < .05$)

No aspects of single task naming performance were significantly correlated with the F*R interaction in dual task naming. However, there were three aspects of dual task naming performance that were significantly correlated with single task naming performance. (1) The frequency effect in dual task naming was significantly correlated with the single task naming frequency effect, the single task naming overall regularity effect and the single task naming low frequency regularity effect. It should be noted that this was the only correlation in Table 5 that was significant using a Bonferroni adjusted p value. (2) The overall regularity effect in dual task naming was significantly correlated with the frequency effect in single task naming. (3) The low frequency regularity effect in dual task naming was significantly correlated with the frequency effect in single task naming.

Measures of Experience and Single Task Naming

The measures of experience that were collected were not found to be strongly related to dual task naming performance. As a further check of the relationship between the experience measures and naming performance, the 4 experience measures used in Table 4 were compared with single task naming performance. These correlations appear in Table 6.

	Experience Measures			
Single Task Measures	Reading	TV	ART	MRT
F*R	0.17	-0.05	-0.13	0.11
Frequency	-0.33	0.33	-0.63 ***	-0.46 *
Overall Regularity	0.07	0.07	-0.27	0.19
LF Regularity	0.13	0.01	-0.21	0.15

Table 6: Pearson correlation matrix for experience measures and single task naming performance. (* $p < .05$; ** $p < .01$, *** Bonferroni adjusted $p < .05$)

None of the measures of experience that were significantly correlated with the frequency by regularity interaction in single task naming. As with the dual task naming condition, experience was related to other aspects of naming performance. Author and magazine recognition test scores were significantly negatively correlated with single task naming frequency effects such that high recognition test scores tended to be paired with smaller frequency effects. The only correlation that was significant with a Bonferroni adjusted p value was between author recognition test scores and the size of the frequency effect.

Discussion

This experiment asked three questions. The first question was whether naming performance in general and the frequency by regularity interaction in particular were significantly different in single and dual task naming. Naming performance in general was significantly slower in dual task naming than in single task naming. It should be noted that the 153 ms. slowdown when going from single task naming (0 item load) to dual task naming (1 item load) is larger than the average 20 ms. slowdown which Bernstein and Carr (1996, Experiment 2) observed when varying memory load from 1 to 5 items. It is unlikely that the attention demand of maintaining a single digit in memory is responsible for the difference between single and dual task naming. A more likely possibility is that the dual task paradigm quantitatively changed naming performance. Naming stimuli in the dual task paradigm were presented at a random delay to avoid planned switching of attention from the memory task to the naming task. The large slowdown from single to dual task naming could be an indication that subjects respond to this random interval by delaying their naming responses.

This finding could potentially change the way in which the effects of dual task demands on naming should be interpreted. The memory load might not only handicap the operation of the assembly of phonology (Paap & Noel, 1991; Bernstein & Carr, 1996) but also delay the attended processing of the naming response. This would in theory give any automated processing of the naming stimulus a head start over attended processing. Dual route models predict that this should reduce the size of the frequency by regularity interaction, because it extends the range of word frequencies (downward) for which retrieved phonology will tend to finish earlier than assembled phonology. That is, giving retrieval a head start in a dual route model could reduce competition for low frequency words. Because this happens in the 1 item load condition of the dual task paradigm it reduces the extent to which a further increase in load from one to five items can cause a further release from competition. There was no support for this hypothesis in the results of this experiment. If this were the case, there would have been an interaction of frequency, regularity, and load found in this experiment.

The results revealed no change whatsoever in the frequency by regularity interaction between single and dual task naming. The interaction of frequency, regularity, and load was not present, $F(1,23) = 0.06$, $p=.81$, $MSe=1506$. There were no significant differences in the pattern of the two panels in Figure 3. The frequency by regularity interaction was also present in the analysis of naming errors. In this analysis as well, the frequency by regularity interaction was unaffected by task. One implication of this finding is that a reliable measure of the frequency by regularity interaction in single task naming could be used as an independent predictor of the size of this interaction in dual task naming -- if a reliable measure were to be found.

The second question was to look for such a reliable measure. To answer this question a comparison was made of the reliability of measures of individual differences in susceptibility to regularity effects in general and in the appearance and magnitude of the frequency by regularity interaction in particular. The continuous measure of the frequency by regularity interaction was marginally reliable in that scores on the variable were nearly significantly correlated within subjects in the split-half analysis. The lack of significance is attributable to the small number of subjects used in this experiment -- power for this correlation was approximately .50 (using tables from Rosenthal & Rosnow, 1991). These tables indicate that approximately 45 subjects would have been required to achieve power of .80.

Unlike the continuous measures, none of the categorical measures of the appearance and/or magnitude of the frequency by regularity interaction were reliable. In the split-half analysis, classifications within subjects were not associated significantly more strongly than chance as assessed with Fisher's exact probability test. The conclusion these results warrant is that the more reliable means of measuring the frequency by regularity interaction is with the continuous variable. One concern raised by these results is that because the reliability of this variable was only moderate, it might be neither an effective index nor predictor of the frequency by regularity interaction in dual task naming.

The third question in this experiment was to assess the effectiveness of different variables as predictors of the frequency by regularity interaction in dual task naming. The magnitude of the frequency by regularity interaction in single task naming was not a good predictor of its magnitude in dual task naming in that the two variables were not correlated. This result contradicted the outcome of the ANOVA, which did not detect any effect of load on the frequency by regularity interaction. The non-significant correlation of the size of the single and dual task frequency by regularity interaction could then be interpreted as an error, attributable to the borderline reliability of the continuous variable that measures the size of the interaction. That is, using a moderately reliable variable to predict performance on a second moderately reliable variable did not work. There were no other variables, including single task naming data and experience measures, which were significantly correlated with the magnitude of the frequency by regularity interaction in dual task naming.

A final consideration based on these results was how to best control for the variation in the frequency by regularity interaction. The reason this was an issue is that the RFC effect, the influence of increasing memory load on the interaction of frequency and regularity, could be masked by individual differences in the magnitude of the frequency by regularity interaction under a low memory load. This variation could be controlled for by removing those subjects who are categorized as not displaying the frequency by regularity interaction (e.g. Bernstein & Carr, 1996). This categorical measure of the presence of the frequency by regularity interaction did not prove to be stable in the split half analysis. Furthermore, the results of the critical analysis of the

Bernstein and Carr (1996) data did not reveal a categorical pattern in the magnitude of this interaction, which should have occurred if this measure were justifiable. Consequently, this method of removing variance could introduce error. The linear regression solution is the alternative having the least possibility of introducing error. For the regression to work, it is necessary to have a variable which measures the magnitude of the frequency by regularity interaction under a 1 item load, or preferably an independent factor that is highly correlated with it. There were no reliable independent predictors of the variation in the F*R interaction under a 1 item load. Therefore, individual differences in the 1 item load F*R Experiment 2 must be measured directly for use in the regression.

CHAPTER 4

Experiment 2

Experiment 2 was a test of the GPC hypothesis. The naming stimuli were a set of high and low frequency GPC regular and GPC exception lexical hermits, which are words which have no word body neighbors. These stimuli were used in the dual task naming paradigm to directly test the hypothesis that the RFC effect is due to competition between an incorrect GPC generated pronunciation and a correct retrieved pronunciation. Because the GPC exception hermits violate GPC rules, dual route GPC models of naming predict that the RFC effect must be present in this experiment. Because lexical hermits have no inconsistent word body neighbors to interfere with their naming, single route analogy models predict that the RFC effect must not be present in this experiment. Single and dual route PDP models could accommodate either outcome because the network can capture regularity at many levels of analysis.

The effect of individual differences in susceptibility to regularity effects was assessed using the /continuous measure of the frequency by regularity interaction. The measure was calculated from two independent sources, (a) the single task naming post-test and (b) the low load condition of the dual task naming condition. These each were used as the individual differences variable in separate General Linear Model (GLM) analyses. The GLM analysis thus could be used to evaluate two possibilities. First is that the magnitude of the frequency by regularity interaction predicts how the interaction will be affected by an increase in memory load. Second is the possibility that individual differences in susceptibility to regularity effects are masking the effects of load.

Author recognition test scores and the magnitude of the frequency by regularity interaction in single task naming did predict small amounts of differences in the magnitude of the frequency by regularity interaction in the dual task naming condition of Experiment 1. These measures were collected in Experiment 2 following the dual task naming condition. A separate GLM analysis was performed using author recognition test scores as the individual differences variable. This analysis served to evaluate two similar possibilities. First is that individual differences in print exposure could predict how the interaction of frequency by regularity is affected in dual task naming. Second is the possibility that individual differences in print exposure mask the effects of load.

Method

Subjects

An additional 30 subjects who did not participate in the first experiment were recruited from the subject pool at Michigan State University and received class credit as compensation for their participation. Participation was restricted to native speakers of English with normal or corrected to normal vision.

Apparatus

The apparatus was identical to Experiment 1.

Materials

Memory load stimuli were identical to those used in Experiment 1. Naming stimuli were the list of 96 lexical hermits which appears in Appendix C. These stimuli consisted of non-productive word rimes in that their onset-rime neighborhood size was 1 as determined by an analysis of CVC words in Kucera and Francis (1967) provided by Jared, McRae, and Seidenberg (1990). The stimuli included an equal number of medium frequency regular (mean frequency = 65), medium frequency exception (mean frequency = 74), low frequency regular (mean frequency = 5) and low frequency exception (mean frequency = 5) words. Regularity was determined according to Venezky's (1970) rules. The words were matched as closely as possible in bigram frequency (Solso, Barbuto, & Juel, 1979). The reading experience and media habits questionnaire and the author and magazine recognition tests used in Experiment 1 were also used in this experiment. The single task naming stimuli were the list of high and low frequency GPC regular and exception words which were also used in Experiment 1.

Procedure

For each subject, the list of naming stimuli was randomized within categories before being combined with the memory load stimuli. In order to avoid artifacts due to a particular word always being presented with a particular set of memory stimuli, words were assigned to memory stimuli in a completely random fashion. Each trial began with the visual presentation of a fixation point for a duration of 1000 ms which was in turn followed by the presentation of one or five randomly selected digits (with replacement) for the memory load task. The times the study set remained on the screen were adjusted for the number of items presented to avoid the items in 1 item sets being studied longer and thus being more familiar to the subjects than the items in the 5 item sets. Items appeared for 400 ms each; that is, the one item set was displayed for 400 ms and the five item set was displayed for 2,000 ms.

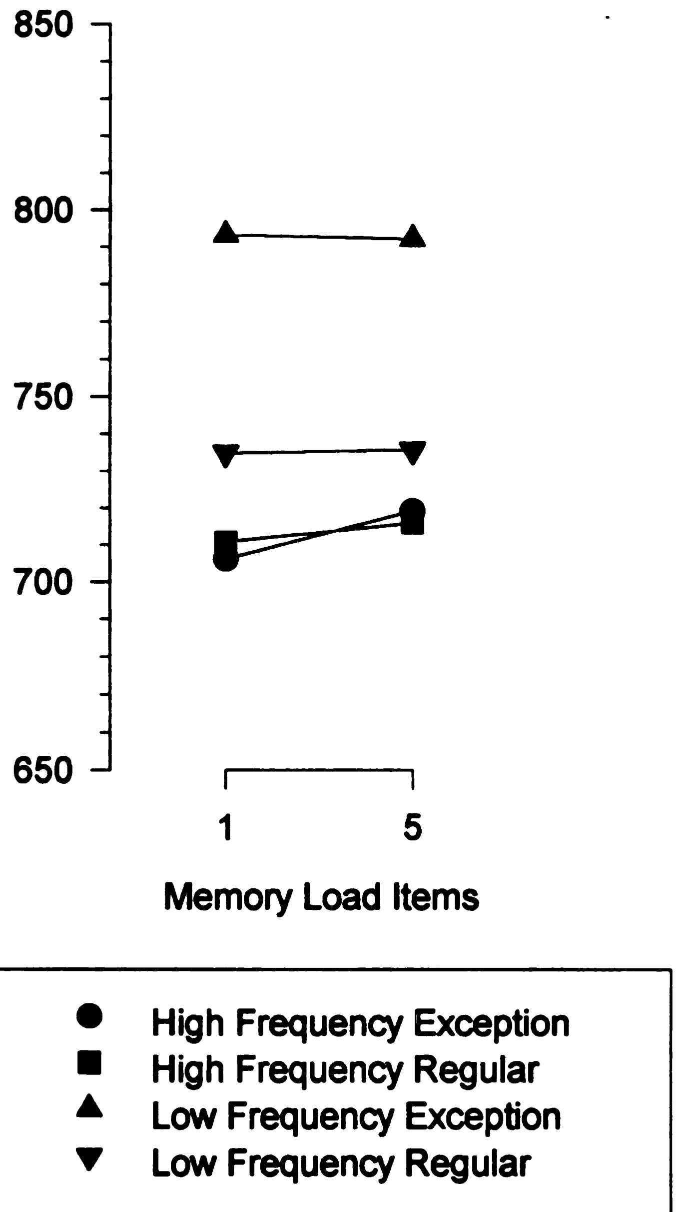
The naming stimulus appeared following a delay which varied randomly between 1 and 2 seconds after the offset of the memory set, during which time a blank screen was displayed. The randomness of the delay was intended to eliminate any planned switching of attention to the naming task. The naming stimulus remained on the screen until the subject responded. After the response a blank screen was displayed until 4 seconds elapsed from the offset of the memory set. This blank screen was followed by the presentation of the memory probe item which remained until the subject responded. Subjects were instructed to press the left mouse button, marked "yes", if the item had appeared in the study set or press the right mouse button, marked "no", if the item had not appeared in the study set. Errors in memory responses resulted in the sounding of a warning tone, followed by a one second presentation of a screen containing the message "error -- please be more careful." An inter-trial interval of 3 seconds elapsed between the memory response, or the offset of the error message, and the beginning of the next trial.

Four practice trials were first completed by the subject, two with a one item load and two with a five item load. These practice trials used memory load stimuli identical to those in the experimental trials and naming stimuli which did not appear in the list of experimental stimuli. A break followed the practice trials, and the subject began the 98 experimental trials at will. The 98 experimental trials began with 2 filler trials which were discarded in the data analysis. The experimental trials took subjects about 25 minutes to complete. Following the completion of the experimental trials, subjects participated in the single task naming condition. Trial timing was identical to Experiment 1. Finally, following the completion of single task naming, subjects were given the reading and media habits questionnaire and the author and magazine recognition tests.

Results

Latencies for correct pronunciations are plotted in Figure 4. Naming latencies were discarded from trials on which there were pronunciation errors (3.2%) or voice key triggering errors due to environmental noises (< 1%). Latencies 125 ms or shorter were also regarded by the experimenter as voice key triggering errors (<1%). Naming latencies longer than 2,000 ms were scored as null responses (< 1%). Pronunciation errors were determined by the experimenter who observed the session and uncertainties regarding naming responses were resolved by listening to the audio tape. Naming latencies were also discarded from trials on which there were memory errors (3%).

Figure 4: Experiment 2 Naming Latencies



Naming latencies were averaged across items and analyzed with subjects as the random factor in an ANOVA with frequency (2), regularity (2), and memory load (2) as within groups factors. Naming latencies were also averaged across subjects and analyzed with items as the random factor in an ANOVA with memory load (2) as a within groups factor and frequency (2), as well as regularity (2) as between groups factors. Naming latencies in this experiment displayed the standard frequency * regularity interaction pattern. Subjects named high frequency words (713 ms) significantly faster than low frequency words (764 ms), $F(1,29)=36.446$, $p < 0.001$, $MSe = 4268$ by subjects; $F(1,92) = 10.311$, $p < .01$, $MSe=12882$ by items. Naming latencies for regular words (724 ms) were significantly faster than exception words (753 ms), $F(1,29) = 24.465$, $p < 0.001$, $MSe = 1976$ by subjects, but not by items, $F(1,92) = 3.229$, $p > .05$. There was a significant interaction of frequency and regularity by subjects, $F(1,29)=24.214$, $p < .001$, $MSe=2091$, but not by items, $F(1,92) = 2.555$, $p > .05$, such that low frequency exception words were named 58 ms slower than low frequency regular words but there was no difference in the naming latencies of high frequency exception and regular words. Increasing memory load from 1 to 5 items had no detectable impact on naming latencies. The main effect of memory load was not significant. The only interaction involving memory load with a $F > 1$ was that of frequency and load, which was significant by items, $F(1,92) = 6.857$, $p < .01$, $MSe = 2043$.

		High Frequency Exception	High Frequency Regular	Low Frequency Exception	Low Frequency Regular
Naming Error Rate	1 item load	0.465	1.208	7.942	1.346
(% errors)	5 item load	0.516	2.717	12.584	2.885
Memory Error Rate	1 item load	3.708	3.902	4.481	3.937
(% errors)	5 Item Load	3.388	3.327	3.263	2.498

Table 7: Experiment 2 Naming & Memory Error Rates

Incidence of naming errors as a proportion of the number of items in each condition appear in Table 7. Naming error rates were analyzed in the same fashion as naming latencies in order to evaluate the possibility of a speed-accuracy tradeoff in naming responses. Error rates were significantly greater for naming low frequency (6.2%) vs. high frequency (1.2%) words, $F(1,29)=53.545$, $p < .001$, $MSe=.003$. Error rates were significantly greater for exception (5.4%) vs. regular (2.0%) words, $F(1,29) = 31.562$, $p < .001$, $MSe=0.002$. There was a significant interaction of frequency and regularity, $F(1,29) = 62.000$, $p < .001$, $MSe=0.002$, such that the effect of regularity on error rates was greater for low frequency (8.2%) than high frequency (1.5%) words. Increasing memory load from 1 to 5 items had no detectable impact on naming error rates. Neither the main effect of memory load nor any interactions involving memory load approached significance (all F's < 1).

Incidence of memory errors as a proportion of the number of items in each condition also appear in Table 6. Memory error rates were analyzed in the same fashion as naming latencies in order to evaluate the possibility of a speed-accuracy tradeoff between the naming and memory tasks. There was no detectable impact of any factor on memory accuracy. The only effect with an $F > 1$ was the non-significant effect of memory load, for which $F(1,26) = 1.260$, $p = 0.271$, $MSe = 0.004$.

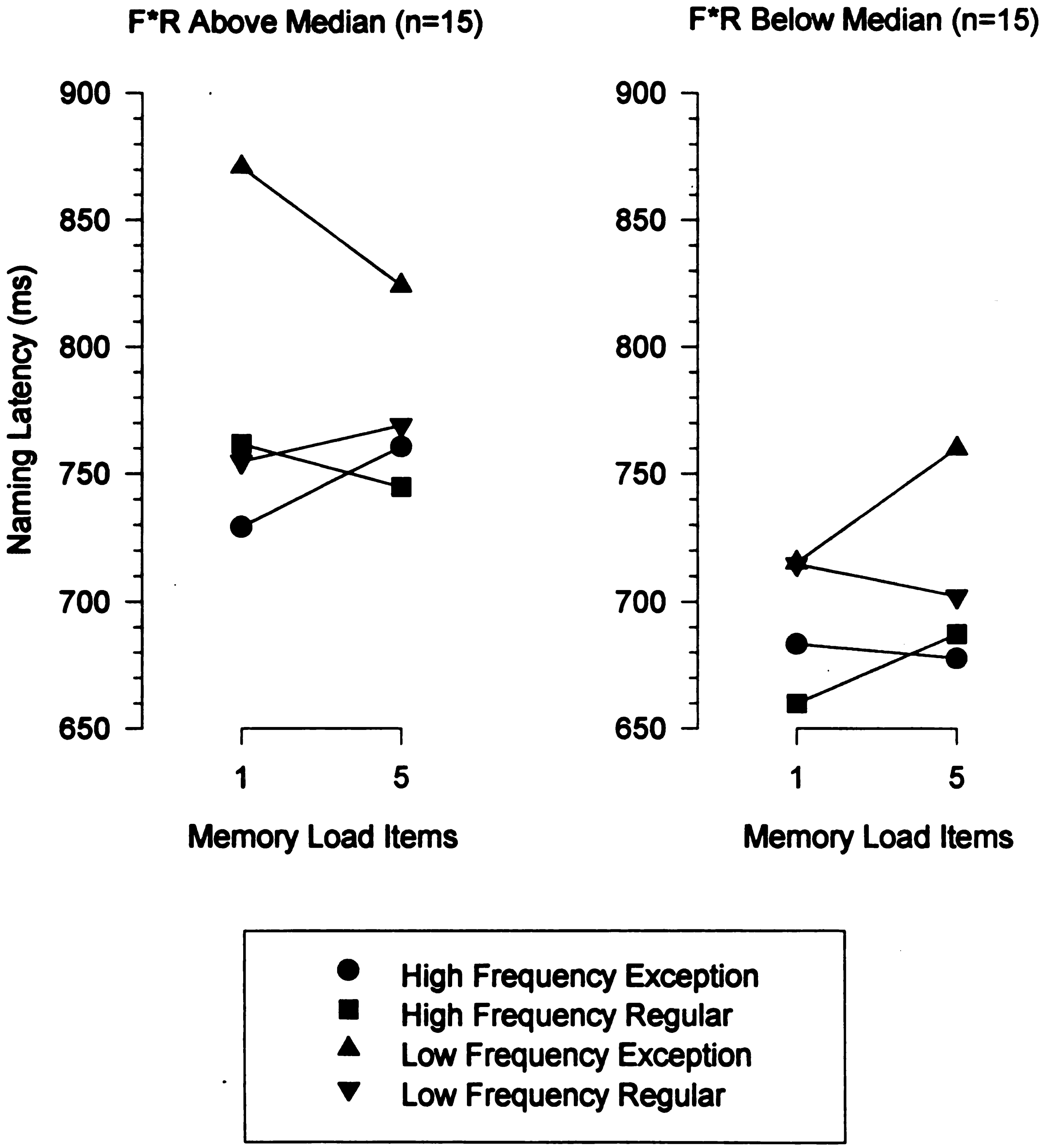
Accounting for Individual Differences

A mixed model regression analysis was employed to evaluate the hypothesis that individual differences in the appearance of the frequency by regularity interaction could be masking the RFC effect. The naming latencies in this analysis were averaged across items with subjects as the random factor. The within groups factors were frequency (2), regularity (2), and memory load (2). A fourth factor was a continuous measure of individual differences. The effects of 3 different individual differences variables were compared in 3 separate analyses. The first individual differences variable was employed for the purpose of comparison with existing data on the RFC effect. It was the continuous measure of the magnitude of the frequency by regularity interaction, based on each subject's 1 item load naming latencies from the dual task paradigm. The second individual differences variable was independent of dual task performance. It was the continuous measure of the magnitude of the frequency by regularity interaction, based on each subject's single task naming post-test data. The third individual differences variable was the score on the author recognition test.

The first regression analysis used the continuous measure of the frequency by regularity interaction from the 1 item load condition of the dual task naming paradigm, REGID1 (**REG**ularity **I**ndividual **D**ifferences at a **1** item load). The magnitude of the frequency by regularity interaction under a 1 item memory load was a significant predictor of the overall magnitude of the frequency by regularity interaction in that there was a significant interaction of REGID1, frequency, and regularity, $F(1,28) = 45.12$, $p < .0001$, $MSe = 829$. The magnitude of the frequency by regularity interaction under a 1 item memory load was also a significant predictor of the interaction of frequency, regularity, and load, $F(1,28) = 68.39$, $p < .0001$, $MSe = 829$. This interaction is graphed in Figure 5, where a median split on the continuous measure of the F*R interaction has been used to roughly illustrate the extremes of the continuous effect. Subjects who displayed large 1 item load frequency by regularity interactions display the F*R*L interaction while subjects with small 1 item load frequency by regularity interactions do not display the interaction. No other interactions involving this factor were significant. The interaction of frequency, regularity, and load was itself significant in the mixed model's report of F's using type III sums of squares. That is, when the effects of the individual differences variable were covaried out of the error variances, the interaction among the means plotted in Figure 4 was significant, $F(1,28) = 12.22$, $p < .01$, $MSe = 829$.

Figure 5: Effect of Individual Differences

In Sensitivity to GPC Regularity



The second regression analysis used the continuous measure of the frequency by regularity interaction from single task naming, REGID0 (**REG**ularity **I**ndividual **D**ifferences at a **0** item load). Unlike the dual task version of this factor, REGID0 did not significantly interact with any of the effects of regularity in dual task naming.[4] The only significant effect of REGID0 was the interaction of frequency and REGID0), $F(1, 28) = 4.66$, $p < .05$, $MSe = 3790$. There was a trend towards a significant main effect of REGID0, $F(1,28) = 3.12$, $p = 0.0883$, $MSe = 70539$. The interaction of frequency, regularity, and load was not significant in this mixed model's report of F's using type III sums of squares ($F < 1$). This factor was based on a difference score which was only marginally reliable in Experiment 1. The null results of the regression analysis are most likely due to this reliability problem.

The third regression used a factor which measures individual differences in print exposure, author recognition test (ART) scores. The three way interaction of ART, frequency, and regularity did not approach significance ($F < 1$), indicating that print exposure was not an effective predictor of the frequency by regularity interaction in dual task naming. The four way interaction of ART, frequency, regularity, and memory load also did not approach significance ($F < 1$), indicating that print exposure also was not an effective predictor of the way in which memory load affects the interaction of frequency and regularity in dual task naming. The interaction of frequency, regularity, and load was also not significant in this mixed model's report of F's using type III sums of squares ($F < 1$).

---

[4] Inspection of the two extremes of the non-significant effect of continuous variation in 0 item load F*R interaction revealed trends consistent with those found using 1 item load. A median split was performed similar to that which was used to create Figure 5. There was a similar but weaker version of the pattern in Figure 5. Subjects with large F*R interactions under 0 item load displayed the RFC effect pattern while those with small F*R interactions under 0 item load did not display the RFC effect pattern.

Discussion

The main question asked in Experiment 2 was whether dual task demands in naming could reduce the interference which causes low frequency GPC exception words to be named more slowly than low frequency GPC regular words. Dual task naming performance for lexical hermits displayed the expected frequency by regularity interaction. For low frequency words, GPC exception words were named more slowly than GPC regular words. Contrary to expectations, overall naming performance was completely unaffected by memory load.

Three separate mixed model regression analyses were performed to investigate how individual differences interact with dual task naming performance, including the effects of load. Individual differences in the frequency by regularity interaction in dual task naming did significantly interact with frequency, regularity and load. Furthermore, removing error variance due to this factor resulted in a significant interaction of frequency, regularity, and memory load. These results support the conclusion that for subjects with relatively large F*R interactions, graphed in the left panel of Figure 5, an increase in memory load releases them from the competition which causes a low frequency GPC irregular word to be named more slowly than a frequency matched GPC regular word.

The other two GLM analyses were not successful in revealing the RFC effect. The GLM analysis with REGID0 as the individual differences variable did not detect that individual differences in the magnitude of the frequency by regularity interaction were in any way related to the way in which memory load affects dual task naming performance. The GLM analysis with ART scores as the individual differences variable did not detect that individual differences in reading experience, as measured by the author recognition test scores, were in any way related to the way in which memory load affects dual task naming performance. Taken together the results of these two analyses suggest that experience factors and single task naming data are insufficient to control for the individual differences which mask the RFC effect. In turn, this sugests that individual differences which emerge in the dual-task environment are the source of the variation in RFC across readers that has been observed in this experient and in Bernstein and Carr (1996).

CHAPTER 5

Experiment 3

The individual differences analyses in Experiments 1 and 2 were aimed at explaining failures to replicate the RFC effect as caused by sampling error in subjects. An alternative explanation of failures to replicate the RFC effect pattern is that they were due to using a measure of spelling-to-sound relationships which imprecisely corresponded to the competence of readers rather than being due to subject factors such as individual differences in susceptibility to regularity and/or differences in memory span. Experiment 3 evaluated the possibility that the weak RFC effect pattern reflects something about the items rather than the subjects.

In most attempts to replicate the RFC effect, spelling-to-sound relationships were defined by GPC regularity. Venezky's (1970) GPC rules are a classification scheme in which a word is identified as regular if its grapheme to phoneme correspondences are those which occur most frequently in a particular context. Venezky defined context as the single grapheme which followed the grapheme involved in a particular spelling-to-sound correspondence. Venezky defined frequency as type frequency, the number of words in the Kucera and Francis (1967) corpus which contained a correspondence, irrespective of the token frequency of these words. Using these rules to classify stimuli is then accepting the assumptions that regularity is (1) type based and (2) limited in its context sensitivity.

Neighborhood consistency analyses (e.g. Jared et al., 1990) classify words by their rimes rather than by individual graphemes. Consistent words have a rime which is pronounced in only one way (e.g. halt, malt, salt). Frequency can be determined either by type frequency or by token frequency. As explained earlier, Jared et al. (1990) compared type and token frequency based neighborhood analyses, derived from a subset of the Kucera and Francis (1967) frequency counts, and found that naming latencies correspond best to token rather than type frequency. The consistency of a word then is most precisely determined by two factors: (a) support for its pronunciation by higher frequency friends and (b) competition with its pronunciation by higher frequency enemies.

In Experiment 3, spelling-to-sound relationships were defined by Jared's token based neighborhood consistency analysis. Naming latencies for inconsistent words with high frequency enemies were compared with naming latencies for consistent words matched in frequency, bigram frequency, and onset in the dual task naming paradigm. These inconsistent words necessarily varied in their GPC regularity, since consistency and regularity are partially overlapping means of categorizing spelling-to-sound relationships. However, the particular set of words for Experiment 3 was chosen to test two predictions which are unique to consistency based accounts of the RFC effect, such as analogy and PDP models: (a) The speedup in naming latencies with increasing load is due to the removal of interference from a higher frequency enemy. (b) The amount of interference which a higher frequency enemy generates can be tempered by the presence of a higher frequency friend. Therefore, the amount of interference suffered under low load and the size of the speedup seen under high load should be smaller for words with high frequency friends. This should happen because words with high frequency friends are not named as slowly as words without them, and therefore don't need as much help to fend off their enemies. The stimuli which are used in this experiment to test this explanation are two categories of inconsistent words with high frequency enemies, along with their matched controls which were selected from Jared et al. (1990, Experiment 2). The properties of these words and predictions for dual task naming are described in the following two paragraphs.

Low Frequency Friends and High Frequency Enemies

Jared et al. (1990) found that among words which have low frequency friends, those words which have high frequency enemies were named significantly more slowly than matched consistent controls. Neighborhood consistency based accounts of naming predict that in dual task naming, latencies for these words should be subject to a speedup in naming latencies with increasing load, both by subjects and by items, while naming latencies for their matched consistent words should slow down with increasing load. According to a GPC rule based account, naming latencies for the matched consistent words should slow down with increasing load, and the speedup in naming latencies for inconsistent words should be weaker by items than by subjects, since only 50% of the words are GPC exceptions.

High Frequency Friends and High Frequency Enemies

Jared et al. (1990) found that among words which have high frequency friends, words with high frequency enemies were named significantly more slowly than their matched consistent control words. The size of this difference was smaller than that found for words which have low frequency friends and high frequency enemies. Neighborhood consistency based accounts of naming predict that in dual task naming, latencies for these words should be subject to a speedup in naming latencies with increasing load, both by subjects and by items, but this speedup should be smaller in magnitude than that found for words with high frequency enemies and high frequency friends. According to a GPC rule based account, the effect should not appear with these stimuli since only 15% of these words are GPC exceptions. Both accounts of naming predict that the matched consistent words should slow with increasing load.

As in Experiment 2, measures of print exposure and a measure of the frequency by regularity interaction in single task naming were collected following the dual task naming condition. Two mixed model regression analyses were performed to establish the relationship between experience measures and the effects of dual task naming. One analysis used author recognition test scores as the individual differences variable. A second analysis used the magnitude of the frequency by regularity interaction from a single task naming post-test for the individual differences variable.

Method

Subjects

An additional 30 subjects who did not participate in the first experiment were recruited from a list of respondents to a newspaper advertisement asking for paid participants in experiments on reading and memory. They were each paid $5.00. Participation was restricted to native speakers of English with normal or corrected to normal vision.

Apparatus

The apparatus was identical to that used in Experiments 1 and 2.

Materials

A list 80 words from Jared et al. (1990, Experiment 2) appear in Appendix D. The list consisted of 4 categories of words: (a) Low frequency words which have high frequency enemies and low frequency friends. (b) Consistent words matched in frequency, bigram frequency, and onset to the words in category a. (c) Low frequency words which have high frequency enemies and high frequency friends. (d) Consistent words matched in frequency, bigram frequency, and onset to the words in category c.

The list of high and low frequency regular and exception words from Experiment 2 were used for the single task naming post-test. While it may have been more appropriate to use a post-test list which varied in frequency and consistency, methodological considerations suggested otherwise. The benefit of using this list is that individual differences in the magnitude of the frequency by regularity interaction were identified by exactly the same criteria in both experiments.

Procedure

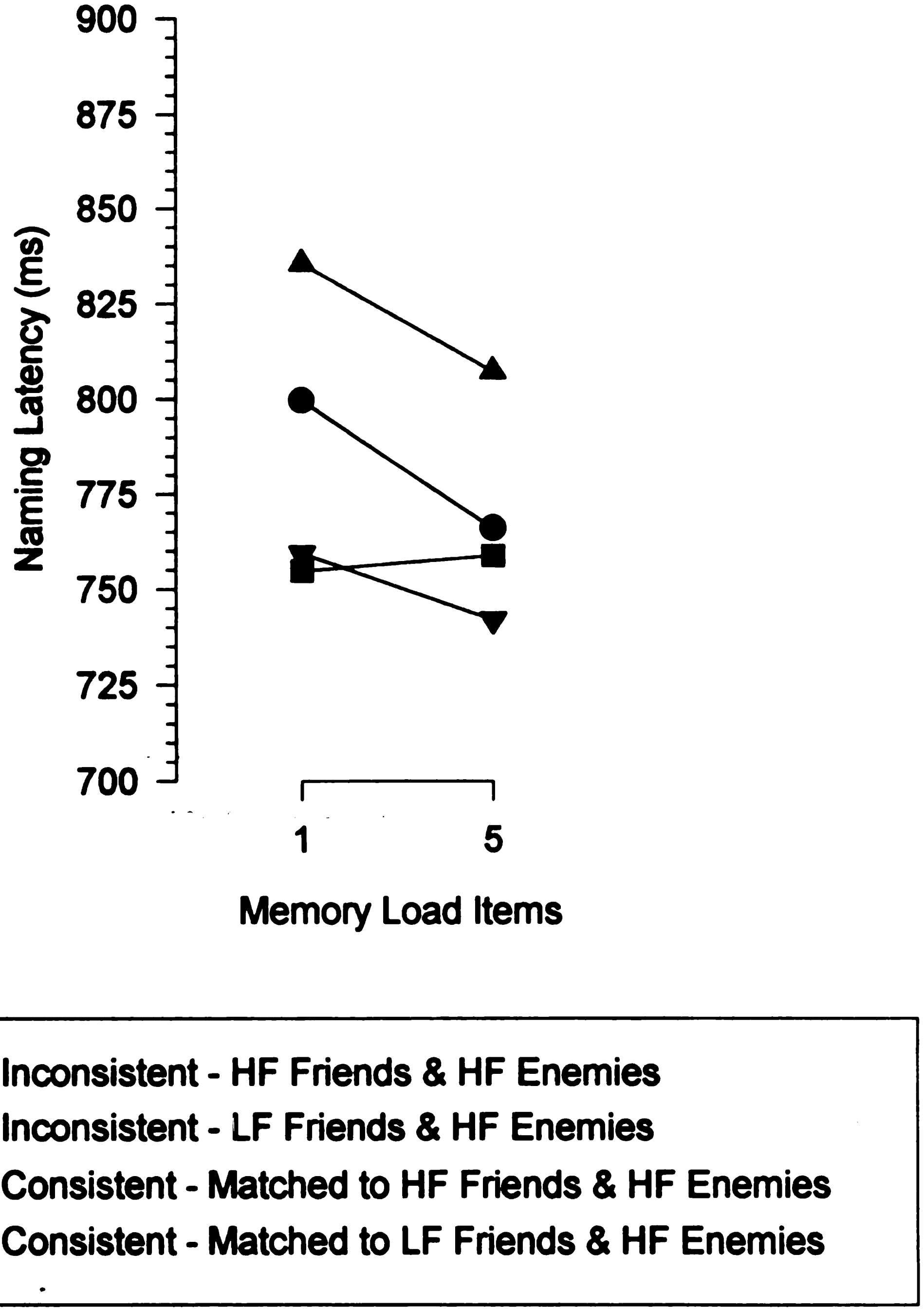
Stimuli were randomized and assigned to conditions as in Experiment 2. The testing procedure and trial timing and structure were identical to Experiment 2.

Results

Naming latencies were discarded from trials on which there were memory errors (3.4%), pronunciation errors (3.2%), or voice key triggering errors due to non-speech noises (1.3%). Latencies 125 ms or shorter were also regarded by the experimenter as voice key triggering errors (< 1%). Naming latencies longer than 2,000 ms were scored as null responses (< 1%).

Pronunciation errors were determined by the experimenter who observed the session and uncertainties regarding naming responses were resolved by listening to the audio tape.

Figure 6: Experiment 3 Naming Latencies



Naming latencies for subjects are plotted in Figure 6. Naming latencies were analyzed in a repeated measures ANOVA with consistency(2), frequency of friends (2), and memory load (2) as within subjects factors. Naming latencies for items were analyzed in a repeated measures ANOVA with memory load (2) as a within groups factor and consistency (2), as well as frequency of friends (2) as between groups factors. Words from inconsistent neighborhoods (803 ms) were named significantly more slowly than words from consistent neighborhoods (754 ms), $F(1,29) = 44.945$, $p < .001$, $MSe = 3133$ by subjects, $F(1,76) = 10.022$, $p < .01$, $MSe = 14029$ by items. There was no significant main effect of having a high frequency friend in that words with high frequency friends (770 ms) were named marginally faster than words with low frequency friends (786 ms), $F(1,29) = 3.489$, $p = .072$, $MSe = 4540$ by subjects, $F(1,76) = 1.593$, $p > .05$, $MSe = 14029$ by items. The interaction of friends and consistency was significant by subjects, $F(1,29) = 9.444$, $p < .01$, $MSe = 3155$ but not by items $F(1,76) = 2.450$ $p =.122$, $MSe = 14029$ such that inconsistent words with a high frequency friend were named significantly faster than those without a high frequency friend.

The overall impact of memory load was to cause a non-significant 19 ms speedup in naming latencies with increasing load, $F(1,29) = 3.271$, $p = .064$, $MSe = 5668$ by subjects, $F< 1$ by items. The interaction of load and consistency was marginally significant by subjects, $F(1,29) = 3.887$ $p = .058$, $MSe = 2290$ but not by items ($F < 1$), such that the speedup in naming latencies for inconsistent items with increasing memory load (31 ms) was greater than the speedup in naming latencies for consistent items with increasing load (7 ms). The three factor interaction of load, friends, and consistency was not significant, $F(1,29) = 0.87$, $p > .05$, $MSe = 3079$.

		Inconsistent HF Friends HF Enemies	Consistent Matched to HF Friends HF Enemies	Inconsistent LF Friends HF Enemies	Consistent Matched to LF Friends HF Enemies
Naming Error Rate	1 item load	2.7	0.7	8. 7	2.7
(% errors)	5 item load	2.7	0.3	8. 7	1. 7
Memory Error Rate	1 item load	4.0	3.7	3.0	4.3
(% errors)	5 Item Load	3.7	2.0	4.0	2.3

Table 8: Experiment 3 Naming & Memory Error Rates

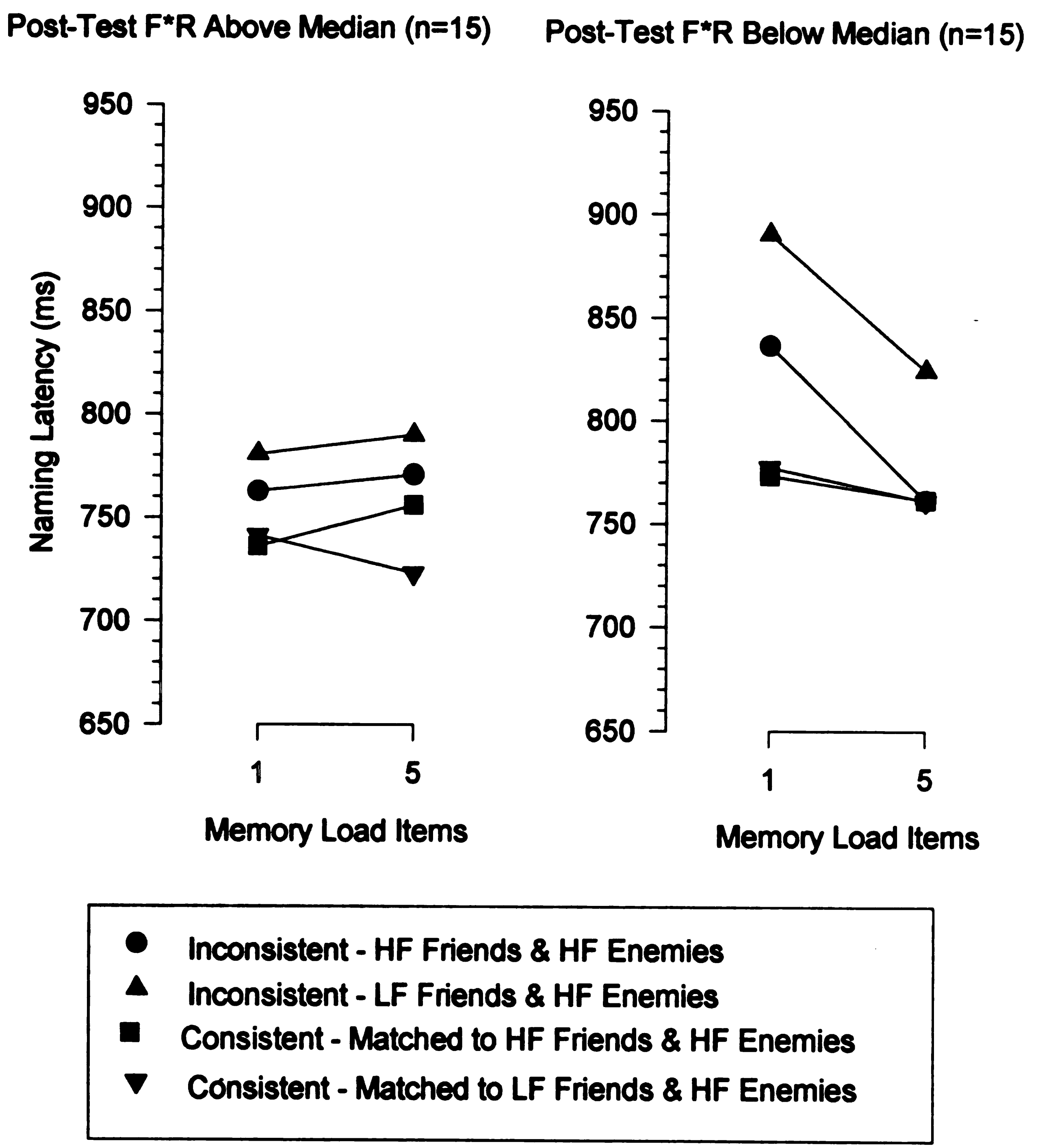
In order to evaluate the possibility of a speed-accuracy tradeoff in the naming part of the dual task paradigm, naming accuracies, which are reported in Table 8, were analyzed (by subjects only) using the same design as the naming latencies. Error rates were significantly greater for words with low frequency friends (0.055) vs. words with high frequency friends (0.016) words, $F(1,29)=20.986$, $p < .001$, $MSe=.004$. Error rates were significantly greater for inconsistent (0.057) vs. consistent (0.014) words, $F(1,29) = 42.250$, $p < .001$, $MSe=0.003$. There was a significant interaction of friends and consistency, $F(1,29) = 11.371$, $p < .001$, $MSe=0.002$, such that the effect of consistency on error rates was greater for words with low frequency friends (0.065) than high frequency friends (0.022) words. Increasing memory load from 1 to 5 items had no detectable impact on naming error rates. Neither the main effect of memory load nor any interactions involving memory load approached significance (all F's < 1).

To evaluate a possible speed-accuracy tradeoff between the memory and naming parts of the dual task paradigm, memory accuracies, which are also reported in Table 8, were analyzed (by subjects only) using the same design as the naming latencies. There were no significant effects of any factor or combination of factors on memory error rates. The only F with a value greater than 1 was the load by consistency interaction, for which $F(1,29) = 2.008$, $p = 0.167$, $MSe = .004$.

Accounting for Individual Differences

As in Experiment 2, a mixed model regression analysis was employed to evaluate the hypothesis that individual differences could determine the appearance of the RFC effect. Two different individual differences variables were used. First, for the purpose of comparison with Experiment 2, subjects were classified using the identical naming post-test. That is, the factor REGID0 was used as the individual differences variable in the GLM analysis. There was a non-significant trend towards individual differences in the magnitude of post-test regularity effects predicting the interaction of consistency and load, $F(1,28) = 1.45$, $p = .2391$, $MSe = 2254$. This trend is graphed in Figure 7, where it can be seen that subjects for whom the continuous measure of the F*R interaction was below median displayed a release from competition for consistent words while subjects for whom the continuous measure was above median displayed a slowdown. Also in this analysis the interaction of consistency and REGID0 was significant, $F(1,28) = 4.28$, $p < .05$, $MSe = 2815$. The interaction of friends and REGID0 was also significant, $F(1,28) = 8.15$, $p < .01$, $MSe = 3641$. Finally, the interaction of load and REGID0 was significant, $F(1,28) = 7.09$, $p < .05$, $MSe = 4685$.

Figure 7: Effect of Individual Differences In Sensitivity to GPC Regularity



The effect of individual differences in the magnitude of the frequency by regularity interaction as measured by the post-test, which appear in Figure 7, was further explored in a different inferential analysis. Subjects were classified by a binary scheme using a median split on the continuous measure of the post-test F*R interaction, creating a between subjects factor with two levels (RC0 -- **R**egularity **C**lassification from **0** item load). This between subjects factor was included in an ANOVA with the within subjects factors Friends (2), Consistency (2) and Load (2) and subjects as the random factor.[5] This analysis revealed that interaction seen in Figure 7 was significant. That is, the magnitude of the *frequency by regularity interaction* in the post-test significantly predicted the presence of the load * consistency interaction, $F(1,28) = 8.55$, $p < .01$, MSe = 1817. The only other significant effect of RC0 was a significant interaction of load and RC0, $F(1,28) = 7.11$, $p < .05$, MSe = 4681.

Accounting for individual differences in print exposure by using author recognition test (ART) scores as an individual differences variable in the GLM analysis did show a trend towards predicting how an increase in memory load would effect the friends * consistency interaction. That is, the four factor interaction of ART with load, friends, and consistency approached significance, $F(1,28) = 3.48$, $p = 0.0725$, MSe = 2836. The interaction of ART, load and friends, a sub-component of this interaction was significant, $F(1,28) = 4.15$, $p = .051$, MSe = 2231. The main effect of load was significant, $F(1,28) = 6.31$, $p < .05$, MSe = 5290, unlike the analysis without the covariate. This indicates than individual differences associated with print exposure are controlled for, naming latencies are significantly effected by load.

---

5 While this measure had proven to be only marginally reliable in Experiment 1, it was assumed to be sufficiently reliable for a crude division of subjects. This median split is not subject to the criticism of regression to the mean, as the Bernstein and Carr (1996) approach was because the split was performed on a predictor variable which was collected independently of the other factors involved in the inferential analysis of the effects of consistency, friends, and load. Similarly, this median split is not subject to the Maxwell & Delaney (1993) criticism, because it was performed on an independent factor.

Discussion

The main question asked in Experiment 3 was whether dual task demands in naming could reduce the interference which causes low frequency inconsistent words to be named more slowly than low frequency consistent words. This was clearly not the case in the overall analysis of the dual task naming condition, where there was a non-significant tendency for the effect of neighborhood consistency to diminish with increasing load. As in previous investigations of the RFC effect, considering individual differences revealed that the overall pattern of results was misleading of the true state of affairs. Subjects who were relatively insensitive to GPC regularity displayed large effects of neighborhood consistency. Furthermore, for these subjects an increase in memory load released them from the competition which an inconsistent word suffers from a higher frequency enemy. This was true regardless of the strength of an inconsistent word's pronunciation in that there was no interaction of load with the frequency of the friends of the inconsistent word. Subjects who were relatively sensitive to GPC regularity displayed small effects of neighborhood consistency. For these subjects an increase in memory load had little effect on naming latencies.

Unlike the previous two experiments, one of the reading experience measures was marginally predictive of dual task naming performance. There was a non-significant trend towards an interaction of author recognition test scores, consistency, and load. This suggests that print exposure may play some role in the individual differences which were detected with the naming post test. Subjects with low author recognition test scores displayed large consistency effects and a release from competition with increasing load. Subjects with high author recognition test scores displayed relatively small neighborhood consistency effects and a small speedup in naming inconsistent words with strong pronunciations, those words with HF friends and HF enemies.

CHAPTER 6

General Discussion

Competition during pronunciation causes spelling-to-sound relationships to influence response latencies in the naming of monosyllabic words. According to the GPC hypothesis, in dual route models of pronunciation, this competition occurs between functionally independent addressed and assembled phonology processes. The result of this competition is an interaction between the frequency and GPC regularity of a word. According to the neighborhood consistency hypothesis, in single route models, this competition occurs within the assembled phonology process, resulting in an interaction between the frequency and neighborhood consistency of a word. These hypotheses were evaluated by examining the influence of two factors that might change the magnitude of this interference. The first factor was individual differences among subjects. The second factor was interference with naming generated by a concurrently performed memory task.

In Experiment 1 variation was found in the magnitude of the F*R interaction that ranged from being absent for some subjects to being relatively large in others. This variation was present in both single and dual task naming, two tasks in which the interaction was not qualitatively different according to the results of the ANOVA. However, the variation in dual task naming performance was not predictable from other factors that included single task naming performance and reading experience. This outcome could be interpreted as an indication that the variation in the magnitude of the F*R interaction in dual task naming is random than reflecting a stable characteristic of subjects. However, the failure to find a measure correlated with the magnitude of the interaction may be explainable by the finding that the variable that was used to measure the interaction itself was only marginally reliable in the split-half analysis. While single task naming performance for the F*R interaction should have been correlated with dual task naming peformance, this outcome was unlikely given that it would require two marginally reliable measures to exhibit a strong amount of covariance. At the end of this experiment it remained the case that the only way to examine the effects of individual differences in the magnitude of the F*R interaction in dual task naming was to measure this interaction directly in the low load condition of the dual task paradigm.

This approach was taken in the second experiment, which was a test of the GPC hypothesis. It was found that subjects with large GPC regularity effects also showed a decrease in the magnitude of these effects with an increase in concurrent task demands. This finding supports the claim that concurrent task interference will cause the RFC effect among those subjects who are sensitive to GPC regularity. A different approach to individual differences was taken in the third experiment, which was a test of the neighborhood consistency hypothesis. The single task naming data, which in Experiments 1 and 2 had been only weakly related to the frequency by GPC regularity interaction in dual task naming, were strongly related to consistency effects in dual task naming. The single task naming data were used to identify those subjects with relatively small GPC regularity effects. These subjects displayed large neighborhood consistency effects and an increase in concurrent task demands caused a decrease in the magnitude of consistency effects. Naming latencies for low frequency inconsistent words became faster with increasing memory load while naming latencies for low frequency consistent words were unaffected by load. In other words, subjects not sensitive to GPC regularity displayed the RFC effect with neighborhood inconsistent words rather than GPC exception words.

Dual Task Naming

The dual task naming paradigm attracted widespread interest because of its potential to cause processing dissociations between the two hypothesized routes from spelling-to-sound in dual route models. Paap and Noel (1991) found both RFC and LoRG in naming latencies. Their results strongly supported the hypothesis that an independent assembled phonology mechanism in a dual route model had been handicapped. This finding was followed by many replication attempts. The RFC effect has proven replicable (Bernstein & Carr, 1996; Herdman et al., in press) but the LoRG effect has only been replicated once, by Herdman et al. (in press) with the digit load. Hypotheses about how parameters of the dual task paradigm might be responsible for failures to replicate were extensively evaluated (e.g. Pexman & Lupker, in press; Herdman et al., in press; Bernstein & Carr, in press). Generally speaking, task manipulations have influenced only the size of the load effects, not the size of the interaction of frequency and spelling-to-sound regularity.

Experiment 1 of this dissertation evaluated a number of potential explanations of why the RFC effect has been difficult to replicate. The first of these reasons was methodological. It was possible that the random delay between the offset of the memory stimuli and the onset of the naming stimuli in the dual task naming paradigm served not only to avoid the planned switching of attention from memory to naming, its intended effect, but also to delay naming responses, an unintended side-effect. The relatively large increase in average naming latencies from single to dual task naming latencies found in Experiment 1 supported the possibility that subjects could have been delaying naming responses. Dual route models predict that delaying naming responses will give retrieved phonology, an automated process, a head start over assembled phonology, an attended process. This would in theory reduce the magnitude of the frequency by regularity interaction, which would not be desirable in the 1 item load condition of the dual task paradigm. The results of Experiment 1 do not support this possibility. The magnitude of the frequency by regularity interaction was completely unaffected by single vs. dual task naming.

The other two potential obstacles to replicating the RFC effect evaluated in Experiment 1 were the subject relevant factor of individual differences and the naming task relevant factor of regularity vs. consistency. It was found that both factors needed to be accounted for to predict how performance will be affected by dual task demands. Individuals varied in the kind of competition that they were subject to, and could subsequently be released from. Some individuals were sensitive to GPC regularity while others were sensitive to neighborhood consistency. The dual task paradigm was able to release subjects only from the particular kind of competition that they exhibited. Subjects sensitive to GPC regularity displayed the RFC effect for stimuli that varied in frequency and GPC regularity. Subjects not sensitive to regularity displayed large consistency effects and these subjects displayed the RFC effect for low frequency inconsistent words. This finding is compatible with the conclusion which Pexman and Lupker (under review) reached about the RFC effect, namely that the important factor in predicting the RFC effect is each individual's ability to handle the particular words used in the experiment.

It is possible that other factors also limit the appearance of the RFC effect. The absence of an overall slowdown with increasing memory load suggests that the dual task demands imposed by the memory task may have been only moderate for the subjects tested in these experiments. One problem which may have caused these weak effects was randomization with replacement for the 5 digit load condition. This could have allowed subjects to recode a 5 digit load into a smaller load. This methodological limitation was unintended. However, the moderate load which was used in these experiments did have significant effects on naming latencies. Furthermore, a post-hoc analysis which was not reported in the results failed to find a graded effect of "true" memory load size. That is, latencies for a 5 item load with 0 repeated items was not significantly slower than a 5 item load with 1 or 2 repeated items. This problem did not prevent replication of the RFC effect in Experiment 2. The only substantive question raised by this flaw is the actual cause of the failure to replicate the LoRG effect in Experiment 2. This failure to replicate the LoRG effect could have been due to not imposing a concurrent task demand high enough to sufficiently handicap assembled phonology, which according to the dual route explanation of the RFC effect is not possible. Handicapping assembled phonology should cause both the RFC and LoRG effects, not one without the other. Future investigations not only should randomize the high load condition without replacement, they also might vary the number of items in the high load condition according to the digit span of each subject. If the LoRG effect is not replicable, the dual route explanation of these results would be called into question.

Individual Differences

Bernstein and Carr (1996) observed that large proportions of subjects in dual task naming experiments failed to display the standard interaction of frequency and regularity in the 1 item memory load condition of the task. Stanovich and West (1988) also observed variations in regularity effects in naming. They found that subjects who scored lower in print exposure were significantly more susceptible to regularity effects in naming than those subjects that scored higher. The present experiments used their measures of reading experience to help determine the cause of individual differences in the magnitude of the frequency by regularity interaction. Individual differences in the magnitude of the F*R pattern were found in all 3 experiments. Experience measures did not account for much of the variation in single and dual task naming. This could have been caused by methodological limitations of the experience measures that were used. The list of authors used in the recognition test was approximately ten years old and might not represent currently popular authors. An updated list should provide a more sensitive measure of experience. The number of subjects employed in the first experiment was rather small for a correlational analysis, which could also explain the weak effects. The lack of an effect of the experience measures could also indicate that there are individual differences in architecture which are not directly related to experience. This possibility is consistent with the finding that experience measures were ineffective at predicting individual differences in susceptibility to the frequency by regularity interaction.

Considerations of individual differences were successful in predicting the conditions that would lead to a replication of the RFC effect. These considerations did not reveal conditions that would lead to a replication of the LoRG effect. Even for subjects who were candidates to posses a dual route architecture, those who displayed a large frequency by regularity interaction, increased memory load did not lead to a relatively large slowdown in the naming of low frequency regular words. This outcome casts doubt on the dual route interpretation of the results of the dual task naming results, namely that assembled phonology has been handicapped. This explanation and others will be evaluated in the next section.

Models of Word Recognition

These experiments were intended to help decide between dual route and single route explanations of the interaction of frequency and the spelling-to-sound relationships in monosyllabic words. No model can explain the results of all 3 experiments. The problem is that the majority of models predict either sensitivity to GPC regularity or neighborhood consistency, but not both. Theoretical frameworks that allow multiple levels of analysis (Glushko, 1979) and those that do not reify any single level (Van Orden et al., 1990) have the potential to accommodate the results of both Experiments 2 and 3. However, these frameworks include no explicit assumptions that can be employed to explain the effects of concurrent task interference. Furthermore, their operating principles are not specified well enough to derive any predictions. A more serious problem is that there is no provision to explain the findings in Experiments 2 and 3 of how these sensitivities vary between individuals.

Dual route models do predict the RFC and LoRG effects and can accommodate variations in sensitivity to GPC regularity. However, these predictions were only partially confirmed in the present experiments. The LoRG effect was not replicated in Experiment 2, which casts doubt on the assembled phonology handicapping explanation. Future replication attempts could include nonwords with the naming stimuli as a more direct test of the hypothesis that the assembly of phonology has been slowed. The RFC effect was replicated in Experiment 2, which supports the idea that competition between an incorrect GPC assembled phonology and a correct retrieved phonology causes low frequency GPC exception words to be named slowly. Dual route models predict a release from this competition when assembled phonology is interfered with more than retrieved phonology. This prediction was confirmed but only for those subjects who displayed relatively large GPC regularity effects. These individual differences can be explained by the flexible coding principle. This principle predicts that competition, and the potential to be released from it, should diminish with experience. This happens because practice causes the speed of retrieved phonology to increase relative to that of assembled phonology. These predictions were not fully supported by the results of Experiment 2. While the magnitude of the regularity effects which subjects displayed was predictive of the RFC effect, no measure of experience was predictive. This result does not directly disprove the flexible coding hypothesis, but neither does it support the hypothesis. More damaging for the dual route theory and flexible coding are the results of Experiment 3. In Experiment 3, small GPC regularity effects were associated with enhanced sensitivity to neighborhood consistency. In a dual route model, an absence of GPC regularity sensitivity is indicative of retrieved phonology, which should be sensitive only to frequency.

Single route PDP models (Plaut & McClelland, 1993; Seidenberg & McClelland, 1989; Van Orden et al., 1990) currently have no operational principles which predict the RFC effect. There is a speculative explanation of the RFC effect from an analogy/PDP standpoint, which might be able to accommodate variations in sensitivity to consistency. These predictions were not fully supported by the results of Experiment 3. In these models, the interference that causes an irregular-inconsistent word to be named slowly is inconsistent crosstalk among orthographically similar but phonologically distinct patterns. This crosstalk pulls an encoding off its attractor, diminishing activation that is needed for the appropriate correspondences to reach threshold. The effect of inconsistent crosstalk is stronger for low frequency words than high frequency words because the strength of an attractor reaches an asymptote when it is high in frequency. Strong attractors are more resistant to the pull. A release from this competition for low frequency words could occur when the memory load in a dual task naming paradigm activates additional attractors, which could dilute the interference which an irregular-inconsistent word suffers from other attractors. The results of Experiment 3 suggest that this can happen without the additional attractors harming other naming latencies. This explanation is tentative in nature. There has been no demonstration that the principles which have been implemented in PDP models could function to achieve this effect.

It is difficult to determine what the effects of individual differences in sensitivity to GPC regularity mean to the single route consistency sensitive models. The problem is that individual differences in sensitivity to GPC regularity were predictive of sensitivity to neighborhood consistency, and RFC for inconsistent words. There was a tendency for those subjects who displayed RFC for inconsistent words to be those who had less print exposure. Therefore, the results of Experiment 3 might not be a problem for PDP models if the individual differences analysis is actually identifying those subjects for whom the frequency norms are inaccurate. It is the results of Experiment 2 that cannot be accommodated within the PDP frameworks, which are sensitive to neighborhood consistency, not GPC regularity.

Conclusion

In summary, both GPC regularity and neighborhood consistency were reliable predictors of interference in naming. Neither regularity nor consistency alone indicated how all subjects responded to all words which they encountered. That is, subjects varied in which measure they were sensitive to. Furthermore, the results of Experiment 2 are inconsistent with the idea that GPC regularity effects are an artifact of failure to control for neighborhood consistency (an idea proposed by D. Jared, personal communication, November 11, 1995). Both kinds of spelling-to-sound relationship effects were reduced by dual task interference, but in different ways for different subjects. For subjects who were relatively sensitive to GPC regularity, an increase in concurrent task demands reduced the interference between an incorrect GPC assembled pronunciation and a correct retrieved pronunciation. For subjects who were relatively insensitive to GPC regularity but were sensitive to neighborhood consistency, an increase in concurrent task demands reduced the interference which an inconsistent word suffered from a higher frequency enemy.

These results could indicate individual differences in system architecture. Some subjects would have dual route GPC architectures. These subjects exhibited a large frequency by regularity interaction for the stimuli used in these experiments. Within the framework of a dual route model this is an indication of interference between retrieved and assembled phonology. Concurrent task interference released subjects from this interference, supporting the dual route account. Other subjects would have single route analogy/PDP architectures. These subjects exhibited a small frequency by regularity interaction for the stimuli used in these experiments. The subjects were sensitive to neighborhood consistency. Within the framework of an analogy/PDP model, this is an indication of interference among simultaneously activated candidates within a neighborhood. Concurrent task interference released these subjects from this kind of interference. This finding cannot be accommodated within the analogy/PDP models. The cause of these apparent differences in system architecture was not revealed in these experiments. While print exposure is a likely candidate on logical grounds, individual differences in naming performance were in no way correlated with measures of print exposure.

Appendices

Appendix A

Reading and Media Habits Questionnaire

Please circle the alternative that is most accurate.

1. I read for pleasure
   a. almost never
   b. a couple of times a year
   c. a couple of times a month
   d. at least once a week
   e. once or more a day

2. Not including textbooks for your college courses, how many books do you read in a year?
   a. none
   b. one or two
   c. 3-10
   d. 10-40
   e. more than 40

3. Excluding the University library, which of the following is true?
   a. I have one library card for a community library.
   b. I do not have a community library card.
   c. I have cards to more than one community library.

4. How many magazines do you yourself (not your family) subscribe to or purchase on a regular basis?
   a. none
   b. one
   c. 2-5
   d. 5-10
   e. more than 10

   Please name as many of these as you can on the following lines.

   ____________________ ____________________ ____________________

   ____________________ ____________________ ____________________

   ____________________ ____________________ ____________________

5. I visit bookstores
   a. never.
   b. once or twice a year.
   c. once or twice a month.
   d. once or more a week.

6. If you answered b or c to the previous question, give the approximate name and town/shopping center of the last bookstore you visited.

   __________________________________ __________________________________

7. If you answered b, c, d, or e to Question 2, please name your two favorite authors/writers.

_______________________ _______________________

8. Regarding newspapers, I usually
   a. read more than one a day.
   b. read a newspaper every day.
   c. read a daily newspaper occasionally.
   d. do not have time to read a daily newspaper.
   e. do not care to read a daily newspaper even if I have the time.

9. On the average, how much television do you watch per day?
   a. I almost never watch television.
   b. Less than one hour.
   c. 1-3 hours.
   d. 3-6 hours.
   e. More than 6 hours.

10. When you were in junior high and high school, on the average how much television did you watch per day?
   a. I almost never watched television.
   b. Less than one hour.
   c. 1-3 hours.
   d. 3-6 hours.
   e. More than 6 hours.

11. In the space below, please list all of the television programs (of all types--e.g., comedy, drama, soap operas, news, sports, cartoons) that you watch on what you would consider a regular basis.

12. Please list the five television programs that were/are your favorites, regardless of whether you watch them regularly and regardless of whether they are still on the air. If there are less than five in this category, please list however many there are.

____

____

____

____

____

Magazine Recognition Test

Below you will see a list of 100 titles. Some of them are the names of actual magazines and some are not. You are to read the names and put a check mark next to the names of those that you know to be magazines. Do not guess, but only check those that you know to be actual magazines. Remember, some of the titles are not those of popular magazines, so guessing can be easily detected.

__ American Journal Review
__ Analog Science Fiction
__ Architectural Digest
__ Architecture Today
__ Atlantic
__ Aviation Week
__ Better Homes & Gardens
__ Business Week
__ Byte
__ Car and Driver
__ Changing Times
__ Commentary Magazine
__ Create
__ Current Health
__ Digital Sound
__ Discover
__ Dow Jones Weekly Report
__ Down Beat
__ Ebony
__ Effervescence
__ Electrical & Mechanical News
__ Elliot
__ Esquire
__ Essence Magazine
__ Family Circle
__ Field & Stream
__ Fitness Today
__ Forbes
__ Future Forecast
__ Galactic Digest
__ Gentlemen's Quarterly
__ Girl Weekly
__ Harper's Magazine
__ Health & Life
__ Home & Yard
__ Home Finance
__ Hot Rod
__ House & Garden
__ Hunters
__ Illustrated Science
__ Industrial Activity
__ Jet
__ Ladies Home Journal
__ Madame
__ Mademoiselle
__ Magellan's
__ Market Trends
__ McCall's Magazine
__ Modern Family
__ Modern Raceway
__ Mother and Child
__ Mother Earth News
__ Mother Jones
__ Motor Sports
__ Motor Trend
__ Mountain and Stream
__ Music Weekly
__ National Geographic
__ Natural History
__ Neuberger Review
__ New Democrat
__ New Yorker
__ New Republic
__ Newsweek
__ Omni
__ Outdoor Times
__ Outdoor Life
__ Pacific World
__ Personal Computing
__ Personal Psychology
__ Planes & Helicopters
__ Popular History
__ Popular Science
__ Progressive
__ Psychology Today
__ Public Policy Review
__ Putnam's American Magazine
__ Reader's Choice
__ Recreation Today
__ Redbook
__ Road & Track
__ Rolling Stone
__ Safeco News Service
__ Science Quest
__ Science Reader
__ Scientific American
__ Software Development
__ Sports Illustrated
__ Stock and Bond Digest
__ Technology Digest
__ Time
__ Tools and Repair
__ Town & Country
__ Trends America
__ Urban Scene
__ Vogue Magazine
__ Wellesley
__ Wellington Home Digest
__ Workbench Magazine
__ World Summary

Author Recognition Test

Below is a list of 100 names. Some of the people in the list are popular writers (of books, magazine articles, and/or newspaper columns), and some are not. You are to read the names and put a check mark next to the names of those individual whom you know to be writers. Do not guess, but only check those whom you know to be writers. Remember, some of the names are people who are not popular writers, so guessing can easily be detected.

__ Maya Angelou
__ Isaac Asimov
__ Jean Auel
__ Pete Axthelm
__ James Baldwin
__ Lawrence Barsalou
__ Derek Besner
__ Judy Blume
__ Kathryn Bock
__ Barbara Taylor Bradford
__ Anthony Burgess
__ Edgar Rice Burroughs
__ Barbara Cartland
__ Arthur C. Clarke
__ James Clavell
__ Charles Clifton, Jr.
__ Jackie Collins
__ Nelson Cowan
__ Fergus Craik
__ Michael Crichton
__ Robert Crowder
__ Len Deighton
__ Susan Duffy
__ Ira Fischler
__ Ian Fleming
__ Dick Francis
__ Nancy Friday
__ John Gardiner
__ Arthur Glenberg
__ Peter Graf
__ Andrew Greely
__ Bob Greene
__ James Hampton
__ Alice Healy
__ Robert Heinlein
__ Frank Herbert
__ Seymour Hersh
__ S.E. Hinton
__ Douglas Hintzman
__ Albrecht Inhoff
__ Larry Jacoby
__ Caren Jones
__ Erica Jong
__ Janice Keenan
__ Stephen King
__ Roberta Klatzky
__ Judith Krantz
__ John K. Kruschke
__ Louis L'Amour
__ Elmore Leonard
__ Doris Lessing
__ Betty Ann Levy
__ Leah Light
__ Gordan Logan
__ Robert Lorch
__ Robert Ludlum
__ Barbara Malt
__ Michael Masson
__ Colleen McCullough
__ Gail McKoon
__ Timothy McNamara
__ Janet Metcalfe
__ James Michener
__ Desmond Morris
__ Toni Morrison
__ James Neely
__ John Nisbitt
__ Robert Nosofsky
__ Laura Novick
__ Lewis Patten
__ Charles Perfetti
__ Matthew Phelps
__ Alexander Pollatsek
__ Sylvia Porter
__ George Potts
__ Jane Bryant Quinn
__ Keith Rayner
__ Arthur Reber
__ Lynn Reder
__ Mike Royko
__ Jay Rueckl
__ Richard Schmidt
__ David Shanks
__ Sidney Sheldon
__ Murray Singer
__ Red Smith
__ Marilyn Smith
__ Danielle Steel
__ Lewis Thomas
__ Alvin Toffler
__ J.R.R. Tolkien
__ Christine A. Vogt
__ Alice Walker
__ Irving Wallace
__ Joseph Wambaugh
__ Edward Wasserman
__ Elke Weber
__ Mary Weldon
__ Garry Wills
__ John Wixted

Appendix B
Experiment 2 Naming Stimuli

LFE	Freq.	Bigram Freq.	LFR	Freq.	Bigram Freq.	HFE	Freq.	Bigram Freq.	HFR	Freq.	Bigram Freq.
bowl	26	7147	beam	20	19624	are	4394	52052	best	360	35710
broad	83	12272	broke	5	10867	both	731	52209	big	370	6717
bush	14	12496	bus	34	12244	break	228	25070	came	622	19328
comb	6	16922	deed	8	29169	choose	50	18348	class	272	17909
deaf	12	18621	dots	11	11452	come	632	25509	dark	193	15889
doll	10	14196	fade	3	16284	do	1373	7135	did	1044	13922
flood	21	8202	float	3	17786	does	485	17801	fact	447	12054
glove	12	13197	grade	49	18201	done	319	31006	flat	77	24120
gross	64	14507	grape	6	13397	door	321	20787	got	482	10532
lose	58	18470	lode	0	15799	foot	113	14557	group	397	18531
pear	6	27308	lunch	33	13691	give	391	14758	him	2619	22096
phase	73	26299	peel	2	17056	good	831	7340	main	120	40222
pint	14	42340	pitch	22	17080	great	670	37868	more	2216	41175
plow	12	12088	pump	14	5268	have	3942	24704	out	2165	25742
pour	9	21789	ripe	14	14823	most	1161	21087	page	73	10570
rouse	2	27300	sank	18	27385	move	171	16433	place	587	16390
sew	6	18030	slam	4	10237	pull	51	12901	see	775	23108
shoe	16	11250	slip	19	9414	put	439	9937	soon	199	27722
spook	0	7880	sock	4	8474	said	1962	11135	stop	122	29568
swamp	5	9384	stunt	1	24570	says	197	7117	tell	271	27938
swarm	3	15940	swore	14	30217	shall	268	27191	wage	70	13054
touch	87	25739	trunk	8	9041	want	328	41375	week	298	9825
wad	0	17167	wake	23	10910	warm	70	20872	well	1004	18485
wand	1	46010	wax	14	9573	watch	85	23311	when	2333	62442
warp	4	19496	weld	1	14567	were	3287	50978	which	3560	22827
wash	35	21948	wick	4	14778	what	1909	34483	will	2686	18069
wool	10	10489	wing	44	46581	word	549	20334	with	7286	62882
worm	8	19940	wit	25	27777	work	496	19209	write	561	27063
mean	21	18444	mean	14	17009	mean	909	23768	mean	1115	24067

Appendix C
Experiment 2 Naming Stimuli - Lexical Hermits

LFE	Freq.	Bigram Freq.	LFR	Freq.	Bigram Freq.	MFE	Freq.	Bigram Freq.	MFR	Freq.	Bigram Freq.
aisle	6	21764	axe	6	681	arc	41	21153	add	89	8829
braille	1	18316	beige	1	12827	court	234	23257	board	239	14019
breadth	7	40858	birch	2	10049	else	176	18349	burst	33	20426
breathe	7	67210	blitz	3	18021	false	29	19601	claim	98	12086
bronze	11	18343	booze	4	4988	film	96	9284	crowd	53	11184
bulge	7	8505	bulb	7	6895	fourth	88	43427	curve	45	13691
choir	8	13782	burnt	7	16657	gauge	16	7090	dealt	22	23399
dirge	2	11050	clothe	1	57975	growth	156	35990	depth	53	42246
dreamt	1	21191	corpse	7	21434	guard	51	13111	desk	65	24451
gourd	3	16427	ebb	1	681	guess	56	16573	doubt	114	11498
hearse	1	46659	gauze	1	2932	guide	36	11302	faith	107	48208
lapse	6	14398	lounge	9	22973	heart	176	49853	fifth	50	38315
mauve	2	13496	mourn	2	18001	laugh	28	9267	glimps	16	12959
pearl	12	21373	scalp	4	15233	meant	100	40340	midst	19	17009
reign	7	20896	sect	2	20233	ninth	23	66880	myth	37	45352
rinse	6	39963	seize	6	10382	realm	19	32394	priest	16	27486
scourge	2	17074	sleeve	11	17542	search	66	24221	sign	95	10416
sheik	4	35804	sparse	5	20515	source	94	18689	sixth	26	38639
sioux	8	19960	strict	11	23285	suite	27	24199	soap	23	7801
sponge	7	25704	taut	9	12963	tongue	40	28897	solve	20	14423
suede	0	21912	tempt	2	16421	twelve	53	13263	staff	116	16650
watt	3	25911	tenth	7	65027	view	187	8872	teeth	103	51444
weird	10	9520	void	10	5812	warmth	28	39150	waist	13	28109
whorl	1	20699	web	6	7188	worst	34	27409	yes	144	24409
mean	5	23784	mean	5	17030	mean	77	25107	mean	67	23460

Appendix D
Experiment 3 Naming Stimuli

Inconsistent HF Friends HF Enemies	Freq.	Bigram Freq.	Consistent	Freq.	Bigram Freq.	**Inconsistent** LF Friends HF Enemies	Freq.	Bigram Freq.	Consistent	Freq.	Bigram Freq.
braid	1	12895	barge	7	15306	beard	26	23488	beech	6	18469
brood	12	11743	broom	2	15115	bough	6	14565	boast	1	21358
crave	2	16358	creep	10	22962	broth	3	44777	brisk	7	17763
flare	3	31081	flail	1	12454	couth	0	52940	carve	3	22250
frost	8	21192	flank	2	23435	foes	6	21489	fern	1	29409
hive	2	23616	hike	4	14725	font	0	37285	fox	12	9346
lone	8	33264	lark	2	19960	lull	2	13240	loin	1	34941
pleat	0	29697	peach	.4	20155	moor	0	22615	mane	0	39334
pose	11	17822	peel	2	17056	mould	0	17795	moist	11	24239
rays	7	11826	reek	2	27157	sneak	2	14452	strict	11	23285
sour	6	22483	seam	9	22646	wove	3	15167	whiff	1	14721
tome	0	29936	tact	6	16191	wreath	8	55697	wretch	1	21039
cleat	1	28684	crust	1	16613	clove	1	13978	cloak	3	6547
hose	9	20232	hiss	1	28923	freak	4	25995	fetch	6	11708
jays	0	4544	jeep	16	7552	hull	11	12673	hoist	1	25884
pare	2	38356	pail	4	13519	moth	2	53768	miff	0	7885
raid	10	15844	reel	3	33271	plough	0	14720	pinch	9	31895
scour	1	21443	sleek	3	12315	sheath	4	69331	screech	1	21282
shave	6	21779	shark	1	22909	woes	0	18363	whisk	3	22315
strive	7	23569	stance	6	33235	wont	2	34158	wane	0	38784
mean	5	21818	mean	4	19775	mean	4	28825	mean	4	22123

Appendix E: Split-Half Analysis -- Categorical Measures

	Split-Half A				Split-Half B			
Subj #	F*R Present Absent	LF Regularity Above Below Median	F*R Present Absent & LF Reg Above Below Median	Continuous F*R Median Split	F*R Present Absent	LF Regularity Above Below Median	F*R Present Absent & LF Reg Above Below Median	Continuous F*R Median Split
1	1	2	3	2	1	1	1	2
2	2	1	2	1	2	2	4	1
3	1	1	1	2	2	1	2	2
4	1	1	1	1	2	2	4	2
5	1	2	3	2	1	1	1	1
6	1	1	1	1	2	2	4	2
7	1	1	1	1	1	1	1	1
8	1	1	1	1	2	2	4	2
9	1	2	3	1	1	2	3	2
10	2	2	4	2	2	2	4	2
11	2	2	4	2	1	1	1	1
12	1	2	3	2	1	1	1	1
13	1	1	1	1	1	1	1	1
14	1	1	1	1	2	2	4	2
15	1	1	1	1	1	2	3	2
16	2	2	4	2	1	2	3	2
17	1	1	1	1	1	1	1	1
18	1	2	3	2	1	1	1	1
19	1	1	1	1	1	1	1	1
20	2	2	4	2	2	2	4	2
21	2	2	4	2	1	2	3	2
22	1	1	1	1	1	1	1	1
23	1	2	3	2	1	2	3	1
24	2	2	4	2	1	1	1	1

Appendix E: Split-Half Analysis -- Continuous Measures

	Split-Half A							Split-Half B						
Subj #	Mean Latency	LF-E Latency	LF Regularity	HF Regularity	F*R	Frequency	Regularity	Mean Latency	LF-E Latency	LF Regularity	HF Regularity	F*R	Frequency	Regularity
1	532	567	25	-1	26	89	24	526	557	38	34	4	50	72
2	580	603	48	-17	65	-3	32	603	607	12	-46	58	-6	-34
3	725	805	68	28	41	183	96	702	733	25	70	-46	75	95
4	613	671	53	-4	57	126	49	604	619	-31	-21	-10	121	-52
5	561	578	19	9	9	29	28	572	612	55	-11	66	48	45
6	603	806	273	-8	281	267	265	574	637	23	26	-4	208	49
7	500	574	56	-11	67	186	45	463	514	53	-16	69	98	37
8	576	617	56	-28	84	52	27	621	670	-42	10	-52	282	-33
9	425	441	27	-24	51	9	3	435	453	5	-8	13	63	-3
10	549	566	12	32	-20	47	44	546	561	-13	-7	-6	85	-20
11	547	558	-2	38	-40	47	37	545	586	68	7	61	31	75
12	573	625	41	1	39	125	42	579	633	60	-13	73	93	48
13	600	750	200	-39	240	199	161	625	723	124	-30	154	144	93
14	593	714	110	-6	116	262	105	615	685	18	24	-6	244	41
15	565	628	94	9	85	63	103	531	548	3	-30	33	59	-27
16	524	572	24	26	-1	142	50	495	531	39	-16	55	68	23
17	599	747	207	32	175	179	239	558	610	99	-23	122	13	76
18	539	550	3	-21	24	41	-17	527	569	61	-12	73	45	48
19	759	864	144	5	139	133	149	769	816	51	-16	67	86	35
20	534	565	30	108	-78	61	138	537	504	-77	171	-248	22	94
21	527	530	14	-8	22	-17	6	525	542	14	-17	31	40	-3
22	732	921	272	-34	306	211	239	783	926	129	54	75	316	182
23	673	708	12	-36	48	114	-24	640	696	23	-114	137	180	-90
24	527	583	-4	-31	27	232	-34	512	598	98	-19	117	148	79

List of References

Andrews, S. (1982). Phonological recoding: Is the regularity effect consistent? Memory & Cognition, 10(6), 565-575.

Balota, D. A., & Chumbley, J. I. (1985). The locus of word-frequency effects in the pronunciation task: Lexical access and/or production? Journal of Memory and Language, 24, 89-106.

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(3), 386-393.

Beauvois, M. F., & Derousene, J. (1979). Phonological alexia: Three dissociations. Journal of Neurology, Neurosurgery and Psychiatry, 42, 1115-1124.

Becker, C. A. (1976). Allocation of attention during visual word recognition. Journal of Experimental Psychology: Human Perception and Performance, 2(4), 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(1), 86-116.

Besner, D. (1984). Specialized processors subserving visual word recognition: Evidence for local control. Canadian Journal of Psychology, 38, 94-101.

Besner, D., Davelaar, E., Alcott, D., & Parry, P. (1984). Wholistic reading of alphabetic print: Evidence from the FDM and the FBI. In L. Henderson (Eds.), Orthographies and reading: Perspectives from cognitive psychology, neuropsychology, and linguistics (pp. 121-135). Hillsdale, NJ: Erlbaum.

Besner, D., & Smith, M. C. (1992). Basic processes in reading: Is the orthographic depth hypothesis sinking? In R. Frost & L. Katz (Eds.), Orthography, Phonology, Morphology, and Meaning (pp. 45-66). Amsterdam: Elsevier.

Besner, D., Twilley, L., McCann, R. S., & Seergobin, K. (1990). On the association between connectionism and data: Are a few words necessary? Psychological Review, 97(3), 432-446.

Bowey, J. A. (1990). Orthographic onsets and rimes as functional units of reading. Memory & Cognition, 18(4), 419-427.

Brooks, L., & Miller, A. (1979). A comparison of explicit and implicit knowledge of an alphabet. In P. A. Kolers, M. E. Wrolstad, & H. Bouma (Eds.), Processing of Visible Language (pp. 391-401). New York: Plenum Press.

Carr, T. H., & Pollatsek, A. (1985). Recognizing printed words: A look at current models. In D. Besner, T. G. Waller, & G. E. Mackinnon (Eds.), Reading Research: Advances In Theory And Practice (pp. 1-82). New York: Academic Press.

Carr, T. H., Davidson, B. J., & Hawkins, J. L. (1978). Perceptual flexibility in word recognition: Strategies affect orthographic computation but not lexical access. Journal of Experimental Psychology: Human Perception and Performance, 4(4), 674-690.

Cohen, J. (1977). Statistical Power Analysis for the Behavioral Sciences (rev. ed.). New York: Academic Press.

Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Eds.), Attention and Performance VI (pp. 535-555). New York: John Wiley & Sons.

Coltheart, M. (1978). Lexical access in simple reading tasks. In G. Underwood (Eds.), 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(4), 589-608.

Coltheart, M., Masterson, J., Byng, S., Prior, M., & Riddoch, J. (1983). Surface dyslexia. Quarterly Journal of Experimental Psychology, 35A, 469-495.

Forster, K. I. (1976). Accessing the mental lexicon. In E. C. J. Walker & R. J. Wales (Eds.), New approaches to language mechanisms (pp. 257-287). Amsterdam: North-Holland.

Funnell, E. (1983). Phonological processes in reading: New evidence from acquired dyslexia. British Journal of Psychology, 74, 159-180.

Gibson, E. J., Pick, A. D., Osser, H., & Hammond, M. (1962). The role of grapheme-phoneme correspondence in the perception of words. American Journal of Psychology, 75, 554-570.

Glushko, R. (1981). Principles for pronouncing print: the psychology of phonography. In A. M. Lesgold & C. A. Perfetti (Eds.), Interactive processes in reading (pp. 61-84). Hillsdale, NJ: Erlbaum.

Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. Journal of Experimental Psychology: Human Perception and Performance, 5(4), 674-691.

Gough, P. B. (1972). One second of reading. In J. F. Kavanagj & I. G. Mattingly (Eds.), Language by ear and by eye. Cambridge, MA: MIT Press.

Gough, P. B., & Cosky, M. J. (1977). One second of reading again. In N. J. Castellan, D. B. Pisoni, & G. R. Potts (Eds.), Cognitive theory Hillsdale, NJ: Erlbaum.

Hawkins, H. L., Reicher, G. M., Rogers, M., & Peterson, L. (1976). Flexible coding in word recognition. Journal of Experimental Psychology: Human Perception and Performance, 2(3), 380-385.

Hays, W.L. (1994). Statistics (5th ed.). New York, NY: Harcourt Brace.

Henderson, J. M., Dixon, P., Petersen, A., Twilley, L. C., & Ferreira, F. (1995). Evidence for the use of phonological representations during transsaccadic word recognition. Journal of Exprimental Psychology: Human Perception and Performance, 21(1), 82-97.

Henderson, L. (1982). Orthography and word recognition in reading. London: Academic Press.

Herdman, C.M. & Beckett (under review). Journal of Experimental Psychology: Human Perception and Performance.

Humphreys, G. W., & Evett, L. J. (1985). Are there independent lexical and nonlexical routes in word processing? An evaluation of the dual-route theory of reading. The Behavioral and Brain Sciences, 8, 689-740.

Jacobs, A. M., & Grainger, J. (1992). Testing a semistochastic variant of the interactive activation model in different word recognition experiments. Journal of Experimental Psychology: Human Perception and Performance, 18(2), 1174-1188.

Jacobs, A. M., & Grainger, J. (1994). Models of visual word recognition -- Sampling the state of the art. Journal of Experimental Psychology: Human Perception and Performance, 20(6), 1311-1334.

Jared, D. (1995). Spelling-sound consistency affects the naming of high-frequency words. Paper presented at the 36th Annual meeting of the Psychonomic Soceity, Los Angeles, CA.

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

Jared, D., & Seidenberg, M. S. (1991). Does word identification proceed from spelling-to-sound to meaning? Journal of Experimental Psychology: General, 120(4), 358-394.

Kay, J., & Bishop, D. (1987). Anatomical differences between nose, palm, and foot, or, the body in question: Further dissection of the processes of sub-lexical spelling-sound translation. In M. Coltheart (Eds.), Attention and performance XII: The psychology of reading Hillsdale, NJ: Erlbaum.

Keppel, G. (1991). Design and analysis: A researcher's handbook (3 ed.). Englewood Cliffs, NJ: Prentice Hall.

Lukatela, G., Lukatela, K., & Turvey, M. T. (1993). Further evidence for phonological constraints of visual lexical access: TOWED primes FROG. Perception & Psychophysics, 53(5), 461-466.

Lukatela, G., & Turvey, M. T. (1991). Phonological access of the lexicon: Evidence from associative priming with pseudohomophones. Journal of Experimental Psychology: Human Perception and Performance, 17(4), 951-966.

Lukatela, G., Turvey, M. T., Feldman, L. B., Carello, C., & Katz, L. (1989). Alphabet priming in bi-alphabetical word perception. Journal of Memory and Language, 28, 237-254.

Marshall, J. C., & Newcombe, F. (1973). Patterns of paralexia: A psycholinguistic approach. Journal of Psycholinguistic Research, 2, 175-199.

Maxwell, S. E., & Delaney, H. D. (1993). Bivariate median splits and spurious statistical significance. Psychological Bulletin, 113(1), 181-190.

Maxwell, S. E., Delaney, H. D., & Dill, C. A. (1984). Another look at ANCOVA versus blocking. Psychological Bulletin, 95(1), 136-147.

McCarthy, R., & Warrington, E. K. (1986). Phonological reading: Phenomena and paradoxes. Cortex, 22, 359-380.

McCarthy, R. A., & Warrington, E. K. (1990). Cognitive Neuropsychology: A clinical introduction. San Diego: Academic Press.

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

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(3), 452-467.

Morrison, F. J. (1995). The nature and consequences of individual differences in early literacy skills. Colloquium presented at Michigan State University, East Lansing, MI, USA. February 2, 1995.

Morton, J. (1969). Interaction of information in word recognition. Psychological Review, 76, 165-178.

Paap, K. R., McDonald, J. E., Schvaneveldt, R. W., & Noel, R. W. (1987). Frequency and pronunceability in visually presented naming and lexical decision tasks. In M. Coltheart (Eds.), Attention and performance XII: The psychology of reading. (pp. 221-243). Hillsdale, NJ: Erlbaum.

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

Paap, K. R., Noel, R. W., & Johansen, L. S. (1992). Dual-route models of print to sound: Red herrings and real horses. In R. Frost & L. Katz (Eds.), Orthography, Phonology, Morphology, and Meaning (pp. 293-318). New York: North Holland.

Patterson, K. E., & Coltheart, V. (1987). Phonological processes in reading: A tutorial review. In M. Coltheart (Eds.), Attention and performance XII: The psychology of reading (pp. 421-448). Hove, England: Erlbaum.

Patterson, K. E., Marshall, J. C., & Coltheart, M. (Ed.). (1985). Surface dyslexia: Neuropsychological and cognitive studies of phonological reading. Hillsdale, NJ: Erlbaum.

Patterson, K.E., Seidenberg, M.S., & McClelland, J.L. (1990). Connections and disconnections: Acquired dyslexia in a computational model of reading processes. In R.G.M. Morris (Ed.) Parallel distributed processing: Implications for psychology and neuroscience. London: Oxford University Press, 131-181.

Pexman, P. M., & Lupker, S. J. (in press). Effects of memory load in a word naming task: Five failures to replicate. Memory & Cognition.

Pexman, P. M., & Lupker, S. J. (under review). Word naming and memory load: Still searching for an individual differences explanation. Journal of Experimental Psychology: Learning, Memory, and Cognition.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Paterson, K. E. (1994). Understanding Normal and Impaired Word Reading: Computational Principles in Quasi-Regular Domains (Technical Report No. PDP.CNS.94.5). Carnegie Mellon University.

Plaut, D. C., & McClelland, J. L. (1993). Generalization with componential attractors: Word and nonword reading in an attractor network. In Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society (pp. 824-829). Hillsdale, NJ: Erlbaum.

Pugh, K. R., Rexler, K., & Katz, L. (1994). Evidence of flexible coding in visual word recognition. Journal of Experimental Psychology: Human Perception and Performance, 20(4), 807-825.

Rosenthal, R., & Rosnow, R. L. (1991). Essentials of behavioral research: Methods and data analysis. (2 ed.). New York: McGraw-Hill.

Rubenstein, H., Lewis, S. S., & Rubenstein, M. A. (1971). Evidence for phonemic recoding i visual word recognition. Journal of Verbal Learning and Verbal Behavior, 10, 645-657.

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

Sears, C. R., Hino, Y., & Lupker, J. (in press). Neighborhood frequency and neighborhood size effects in word recognition. Journal of Experimental Psychology: Human Perception and Performance.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. Psychological Review, 96(4), 523-568.

Seidenberg, M. S., Plaut, D. C., Petersen, A. S., McClelland, J. L., & McRae, K. (1994). Nonword pronunciation and models of word recognition. Journal of Experimental Psychology: Human Perception and Performance, 20(6), 1177-1196.

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

Solso, R. L., Barbuto, P. F., & Juel, C. L. (1979). Bigram and trigram frequencies and versatilities in the English language. Behavior Research Methods & Instrumentation, 11(5), 475-484.

Stanhope, N., & Parkin, A. J. (1987). Further explorations of the consistency effect in word and nonword pronunciation. Memory & Cognition, 15(2), 169-179.

Stanovich, K. E., & Bauer, K. W. (1978). Experiments on the spelling-to-sound regularity effect in word recognition. Memory & Cognition, 6, 410-415.

Stanovich, K. E., & West, R. F. (1989). Exposure to print and orthographic processing. Reading Research Quarterly, 24(4), 402-433.

Sternberg, R. J. (1966). High-speed scanning in human memory. Science, 153, 652-654.

Stone, G. O., & Van Orden, G. C. (1994). Building a resonance framework for word recognition using design and system principles. Journal of Experimental Psychology: Human Perception and Performance, 20(6), 1248-1268.

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

Temple, C. M., & Marshall, J. C. (1983). A case study of developmental phonological dyslexia. British Journal of Psychology, 74, 517-533.

Treiman, R., & Chafetz, J. (1987). Are there onset- and rime- like units in printed words? In M. Coltheart (Eds.), Attention and Performance XII: The psychology of reading. (pp. 281-298). Hillsdale, NJ: Erlbaum.

Treiman, R., Mullennix, J., Bijeljac-Babic, R., & Richmond-Welty, E. D. (1995). The special role of rimes in the description, use, and acquisition of English orthography. Journal of Experimental Psychology: General, 124(2), 107-136.

Treiman, R., & Zukowski, A. (1988). Units in reading and spelling. Journal of Memory and Language, 27, 466-476.

Treiman, R., & Zukowski, A. (1990). Toward an understanding of English syllabification. Journal of Memory and Language, 29, 66-85.

Van Orden, G. C. (1987). A ROWS is a ROSE: Spelling, sound, and reading. Memory & Cognition, 15(3), 181-198.

Van Orden, G. C., Johnston, J. C., & Hale, B. L. (1988). Word identification in reading proceeds from spelling-to-sound to meaning. Journal of Experimental Psychology: Learning, Memory, and Cognition, 14(3), 371-186.

Venezky, R. L. (1970). The Structure of English Orthography. The Hague: Mouton.

Waters, G. S., Seidenberg, M. S., & Bruck, M. (1984). Children's and adults' use of spelling-sound information in three reading tasks. Memory & Cognition, 12(3), 293-305.

Wijk, A. (1966). Rules of pronunciation for the English language. London: Oxford University Press.