

Relating distinctive orthographic and phonological processes to episodic memory performance

MICHAEL J. CORTESE

College of Charleston, Charleston, South Carolina

JASON M. WATSON

Washington University, St. Louis, Missouri

JING WANG

Bowling Green State University, Bowling Green, Ohio

and

APRIL FUGETT

Morehead State University, Morehead, Kentucky

Three experiments examined the role of three distinctive perceptual factors in recognition and recall memory. Using a subject-paced presentation rate, the first two experiments (recognition and recall) examined (1) the number of phonological-to-orthographic neighbors, (2) phonological-to-orthographic consistency, and (3) orthographic-to-phonological consistency. The third experiment (recall) reexamined the number of phonological-to-orthographic neighbors, using an experimenter-paced presentation rate of 2 sec per item. In both recognition and recall memory tasks, the number of phonological-to-orthographic neighbors influenced memory performance, whereas the two types of consistency did not. The results indicate that having fewer phonological-to-orthographic neighbors (i.e., having distinct mappings between orthography and phonology, and between phonology and orthography, e.g., *pulp*) relieve words from interference in episodic memory tests for such words. Furthermore, words that are indistinct in terms of these mappings (e.g., *tuck*) are subject to interference from words with similar representations (e.g., *luck*, *buck*, *stuck*), and this weakens the memory trace for a particular word.

The concept of distinctiveness suggests that memory for a given item will benefit to the extent to which that item is unique. For example, the von Restorff effect (see, e.g., Wallace, 1965) refers to better memory performance for items that are made “different” from others during encoding (e.g., by underlining them, presenting them in a different color, etc.). In addition, Hunt and Elliott (1980) demonstrated that items with distinctive word forms (e.g., *phlegm*) were recalled at a higher rate than were items not orthographically distinctive (e.g., *primate*). More recently, Hirshman and Jackson (1997) reported that words with inconsistent orthographic-to-phonological mappings (e.g., *plaid*) were recalled at a higher rate than were words with more consistent orthographic-to-phonological mappings (e.g., *plump*). In these studies, the unique attributes differentiated particular items from others during encoding, and they may have been reactivated during retrieval to facilitate memory for them.

However, as a theoretical construct, distinctiveness can seem to be circular (see Schmidt’s, 1991, discussion); it has been addressed both as an empirical phenomenon and as an explanatory tool. That is, distinctive items are remembered well, and items remembered well are considered distinctive. Our view is that models of word recognition can help show what might make one class of items more distinctive than another. Although these models have not been designed to explain memory performance, they provide a useful framework. For example, in the triangle model (e.g., Plaut, McClelland, Seidenberg, & Patterson, 1996), distributed representations for words exist on three levels: (1) orthography, (2) phonology, and (3) semantics. The same grapheme and phoneme units are used to process all words, but each unique word is associated with a unique pattern of activation among the units. Thus *tuck*, *buck*, and *duck* share two of three grapheme and phoneme units, and would likely interfere with each other more than words sharing fewer units. In localist models (e.g., that of Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001), words in the lexicon will be inhibited more by similar words than by dissimilar words. In both the triangle model and localist models, connections exist between orthography and phonology, both of

We thank Kendra Noble and Clay Barber for their help in collecting and coding data. Correspondence concerning this article should be addressed to M. J. Cortese, Department of Psychology, College of Charleston, 66 George Street, Charleston, SC 29424 (e-mail: cortesem@cofc.edu).

which are likely to become activated when a word is presented. Thus, within this framework, the circularity problem is eliminated because distinctiveness effects arise naturally as a consequence of the models' architecture.

The present study examines the hypothesis that a word is distinct to the extent that it has a unique orthographic and phonological representation in memory and that words that share orthographies and phonologies with a target should produce interference. The interference should weaken the memory trace (or render the memory trace less accessible) for that target word. To test our hypothesis, in three experiments we examined memory for monosyllabic words that share their orthography and phonology with either a small number of words or a large number of words. We focused on the rime unit (i.e., the vowel and subsequent consonants).

Three factors were examined. The *number of phonological-to-orthographic neighbors* was the number of words that shared both the orthographic and phonological rimes with a target word. For example, *tuck* has the neighbors *duck*, *luck*, *puck*, *truck*, *stuck*, and so forth. *Phonological-to-orthographic consistency* was the degree to which similar-sounding words are spelled similarly to a target word. For example, *tuck* is consistent because all of its phonological neighbors are spelled the same. In contrast, *hype* is inconsistent because *ripe*, *gripe*, *pipe*, and so forth share the same sound (e.g., /aip/), but the spelling for that sound is usually different from that found in *hype*. *Orthographic-to-phonological consistency* referred to the degree to which similarly spelled words are pronounced similarly (e.g., Cortese & Simpson, 2000). For example, *plaid* is considered inconsistent because the letter string *aid* is not usually pronounced /æd/ (e.g., *maid*, *raid*, *paid*, *said*, etc.).

We use the terms *orthographic-to-phonological consistent* (OPC) and *inconsistent* (OPI) to describe the orthographic-to-phonological consistency of items. We use the terms *phonological-to-orthographic consistent* (POC) and *inconsistent* (POI) to describe the phonological-to-orthographic consistency of the items. Words with few rhyming neighbors that are spelled the same are denoted as *low-N*, and words with many rhyming neighbors that are spelled the same are denoted as *high-N*. These terms are used to denote the four types of words used in the present study: (1) OPC and POC–low-N (e.g., *pulp*), (2) OPC and POC–high-N (e.g., *tuck*), (3) OPC and POI–low-N (e.g., *hype*), and (4) OPI and POI–low-N (e.g., *plaid*). The different conditions of words were equated for frequency and length. Orthographic neighborhood size (Coltheart, Davelaar, Jonasson, & Besner, 1977) was controlled via analysis of covariance (ANCOVA). These four word types allow for three separate comparisons: (A) The number of PO neighbors (Condition 1 vs. Condition 2), (B) PO consistency (Condition 1 vs. Condition 3), and (C) OP consistency (Condition 3 vs. Condition 4). Comparison A holds OP and PO consistency constant, and the number of PO neighbors varies. Comparison B holds the number of

PO neighbors and OP consistency constant, and PO consistency varies. Comparison C holds the number of PO neighbors and PO consistency constant, and OP consistency (e.g., Hirshman & Jackson, 1997) varies.

We predicted that POC–low-N words would be remembered better than POC–high-N words (Comparison A). Specifically, if phonological-to-orthographic representations are activated during encoding and/or retrieval, then low-N words should experience less interference from their neighbors than should high-N words. In Comparison B, POC–low-N words should not receive interference for reasons already discussed. In contrast, POI–low-N words could receive interference because they share phonologies with a large number of words (e.g., *hype* shares /aip/ with *pipe*, *gripe*, *ripe*, etc.). However, the distinct phonological-to-orthographic mapping of these words might relieve them from interference if interference requires that a target word share both the phonology and orthography with a large number of words. Finally, one might not expect an effect of OP consistency (Comparison C). Both types of words might receive interference from rhyming neighbors. In addition, OPI words might receive interference at the orthographic level (e.g., *plaid* shares a common spelling with *laid*, *aid*, *raid*, etc.). However, if orthographic and phonological overlap is necessary for interference to occur, then neither condition should receive much interference.

EXPERIMENT 1 Recognition Memory

Recognition memory provides the most straightforward way of testing for the influence of the factors listed above. Specifically, many items are difficult to spell owing to their phonological-to-orthographic consistency (see Cortese, Balota, Sergent-Marshall, & Buckner, 2003). Thus, if recall consists of writing down the items, then it may be difficult to ascertain the intent of the response when the items are misspelled.

Another reason for employing a recognition memory test is to isolate as much as possible factors that influence encoding. Because recognition memory minimizes retrieval operations, a factor that exerts its influence in recognition may be important during encoding. However, if retrieval operations provide the mechanism for which the distinctiveness of these factors operates, then one might not expect any of the comparisons to be significant.

Method

Participants. Thirty-four Morehead State University undergraduate psychology students participated in the experiment for extra credit.

Materials. The items used in the study, along with their recall and recognition values, are presented in the Appendix. Stimulus characteristics are presented in Table 1.¹ The items consisted of 56 monosyllabic words of low frequency (Kučera & Francis, 1967) and 20 filler items. Fourteen words were selected from each of four categories: (1) OPC and POC–low-N (e.g., *pulp*), (2) OPC and POC–high-

Table 1
Stimulus Characteristics for the Words Used in the Study

	Frequency	Length	PO Friends	PO Enemies
OPC and POC–Low-N	5.6	4.5	0.5	0.0
OPC and POC–High-N	6.3	4.5	10.7	0.0
OPC and POI–Low-N	7.3	4.3	0.3	12.4
OPI and POI–Low-N	6.0	4.6	0.4	12.6

Note—O, orthographic; P, phonological; C, consistent; I, inconsistent; low-N, few PO neighbors; high-N, many PO neighbors. PO friends refer to the number of words that contain both the same phonological rime and spelling for that rime, and PO enemies refer to the number of words that contain the same phonological rime and a different spelling for that rime.

N (e.g., *tuck*), (3) OPC and POI–low-N (e.g., *hype*), and (4) OPI and POI–low-N (e.g., *plaid*). No two words contained both the same orthographic and the same phonological rime segment.² Between conditions, items were equated for frequency and length. From the original 56 words and 20 filler items, two lists of 38 items were constructed (each containing 28 critical items and 10 filler items). Each list consisted of 7 words from each condition and 10 filler words. In the experiment, each participant studied one of the two lists of 38 items and was tested with one list of 76 items (both lists combined). The study lists were counterbalanced across participants so that each item served equally often as an old and new item in the recognition memory test. PO consistency (i.e., the number of PO “friends” and “enemies” in Table 1) was determined according to the Ziegler, Stone, and Jacobs (1997) norms. The term *PO friends* refers to the number of words that contain both the same phonological rime and spelling for that rime, and the term *PO enemies* refers to the number of words that contain the same phonological rime and a different spelling for that rime. OP consistency was also determined according to the Ziegler et al. norms and/or the rule-based pronunciation of the Coltheart et al. (2001) model. Some words that are classified as consistent by the Ziegler et al. norms because they do not contain any word–body (i.e., orthographic rime) neighbors (e.g., *yacht*, *niche*, *gauge*, *tongue*, *ache*, and *crepe*) were classified as inconsistent in the present study because they are highly inconsistent at the grapheme-to-phoneme level (see Berndt, Reggia, & Mitchum, 1987). We submitted these words to the grapheme-to-phoneme (rule-based) translation procedure of Coltheart et al. (2001), available at <http://www.macqs.mq.edu.au/Utilities/regcelex.html>. This procedure assigns pronunciation codes for words according to the most common grapheme-to-phoneme mappings for monosyllabic words. Indeed, all 6 words were assigned incorrect pronunciation codes.

Procedure. During the study phase, words were presented in random order, one at a time in lowercase letters in the center of the computer screen. Each word appeared until the participant pressed the space bar. The participants were instructed that they would be tested for their memory of the words presented and that they controlled the duration of the stimulus presentation, but that they should not spend too much time on any one word. After the study phase, the participants performed a simple math verification task. Eighteen problems (e.g., $4 + 3 = 6$), with either correct or incorrect solutions, were presented one at a time. The participants were instructed to press the “/” key if the solution was correct and the “z” key if the solution was incorrect. The math task was followed by the recognition memory task. The participants were informed that they would be presented with a list of words, one at a time, half of which had been seen earlier (i.e., *old* items), and half of which had not been seen earlier (i.e., *new* items). They were instructed to press the “/” key if they recognized the item as old and the “z” key if they recognized the item as new. Although decision time was recorded, the participants were not informed of this, nor were they instructed to respond quickly. Therefore, the retention interval was equal to the amount of time it took to read the instructions for the math task,

solve the 18 math problems, and read the instructions for the recognition memory test.

Design. Stimulus type was manipulated within subjects and consisted of four levels (OPC and POC–low-N, OPC and POC–high-N, OPC and POI–low-N, and OPI and POI–low-N). The dependent variables were hits, false alarms, hits – false alarms, encoding time, and decision time (for hits and correct rejections; i.e., correct responses).

Results and Discussion

The results of Experiment 1 are presented in Table 2. It is clear from Table 2 and from the analyses to follow that the only factor that influenced recognition memory was the number of PO friends. Analysis of variance (ANOVA) by subjects (F_s) was performed separately on hits, false alarms, hits – false alarms, encoding time, and decision time. Additional item analyses were conducted (F_i , Clark, 1973) whenever subject analyses yielded significant effects. In the analyses of items, orthographic neighborhood size (Coltheart et al., 1977) was included as a covariate. All effects are reported to be significant at $p < .05$, unless otherwise noted.

In the analysis on hits, the effect of stimulus type was significant for both subjects [$F_s(3,99) = 7.13$, $MS_e = .02$] and items [$F_i(3,51) = 3.53$, $MS_e = .02$]. Additional analyses were conducted for the individual comparisons discussed in the introduction.

In the first set of analyses, the number of PO friends was assessed by comparing OPC and POC–low-N words to OPC and POC–high-N words. The results from these analyses indicated that low-N words had higher hit rates than did high-N words for both subjects [$t_s(33) = 3.22$, $SEM = .04$] and items [$F_i(1,25) = 7.31$, $MS_e = .02$]. Individual comparisons involving orthographic-to-phonological consistency and phonological-to-orthographic consistency were not significant (all t s and F s < 1).

In the analysis on false alarms, the effect of stimulus type was not significant ($p > .23$). Therefore, individual comparisons were not conducted.

In the analysis on hits – false alarms (i.e., adjusted memory), the effect of stimulus type was significant for subjects [$F_s(3,99) = 6.35$, $MS_e = .04$] and marginally significant by items [$F_i(3,51) = 2.66$, $MS_e = .03$, $p < .06$]. The results from the individual comparisons indicated that the adjusted memory for OPC and POC–low-

Table 2
Results From Experiment 1

	Hits		FAs		Hits – FAs		Encoding RT		Decision RT	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
OPC and POC–Low-N	.83	.03	.20	.03	.63	.04	1705.4	206.7	941.0	73.1
OPC and POC–High-N	.69	.03	.22	.03	.47	.05	1752.2	194.2	920.8	53.1
OPC and POI–Low-N	.82	.02	.15	.03	.67	.03	1807.2	232.1	879.9	52.1
OPI and POI–Low-N	.84	.02	.19	.03	.65	.03	1731.9	204.1	905.6	52.8

Note—FAs, false alarms; RT, reaction time (in msec).

N words was higher than that for OPC and POC–high-N words [$t_s(33) = 2.57$, $SEM = .06$; $F_1(1,25) = 5.47$, $MS_e = .03$]. No other comparisons were significant (both t s and F s < 1).

In the analyses on encoding time and decision time, no effects were significant (all F s < 1).

The results of Experiment 1 demonstrate that recognition memory performance for low-frequency words benefits when they contain a low number of PO friends. Interestingly, three conditions of words had few friends, and recognition performance for these words was virtually the same. Neither phonological-to-orthographic nor orthographic-to-phonological consistency produced independent effects. This suggests that the interference generated is not at the phonological level alone, but involves coactivation at the orthographic and phonological levels. In other words, a target word must share both a common phonology and orthography with a large number of words in order for sufficient interference to occur.

Furthermore, these results cannot be explained in terms of the amount of encoding time spent on each item type. Although participants determined the amount of time that they needed to encode each item, they used about an equal amount of time for each word type.

EXPERIMENTS 2A, 2B, AND 2C

Recall Memory

In the following set of experiments, the same comparisons were made as in Experiment 1. However, each comparison was conducted as a separate experiment (e.g., Comparison A = Experiment 2A, Comparison B = Experiment 2B, and Comparison C = Experiment 2C). According to our interference framework, the predictions are essentially the same as in Experiment 1. In Experiments 2A, 2B, and 2C, participants encoded two separate lists of 14 items. In one condition, the two types of stimuli were blocked by stimulus type, and in the other condition, they were mixed. This manipulation allowed us to test whether any distinctiveness effect obtained is list specific or is more general. If the effect is list specific, then we should find that it occurs more strongly when the items are mixed than when they are not. If the effect is general, then the list-type manipulation should not have an effect. We predicted that the list-type manipulation would not be significant because the interference would not be generated by the other items on the list, but by representations stored in memory.

The results of Experiment 1 suggested that the number of PO neighbors has at least a partial basis in encoding. If this is so, we should find that the effect is maintained in a free-recall task as well (Experiment 2A). Although our participants wrote down their responses, we did not foresee a problem with spelling accuracy (in Experiment 2A), as Cortese et al. (2003) found that both of these types of words were spelled with a high degree of accuracy (approximately 98% accurate). We note that spelling accuracy should not be a problem in Experiment 2C. (Cortese et al. 2003, found no difference in the spelling accuracy for these two types of words; i.e., both were commonly misspelled, but at the same rate.³)

Experiment 1 failed to produce an effect of either PO or OP consistency. However, it is possible that one effect or the other has a basis in retrieval operations. In fact, Hirshman and Jackson (1997) have suggested that the basis of the OP consistency effect arises during retrieval operations. Specifically, they claim that during retrieval, OP representations are activated and serve as retrieval cues. Inconsistent words, having more distinct OP mappings, are facilitated more by this process than are consistent words. If so, we should replicate their pattern of results in Experiment 2C, and we might expect that a PO consistency effect would also arise during retrieval if a similar PO process was implemented during retrieval.

Method

Participants. One hundred forty-four Morehead State University undergraduate psychology students participated in Experiments 2A, 2B, and 2C (48 participated in each experiment) for extra credit.

Materials. The words from Experiment 1 were used as stimuli. In each experiment, two experimental lists of 14 pure stimuli, and two 14-item lists of mixed stimuli (7 of each type) were constructed. In both the pure and mixed conditions, lists were counterbalanced across participants so that the first and second lists occurred equally often. A practice list of 20 items (the filler items used in Experiment 1) was also constructed.

Procedure. The participants were presented with three lists of items. The first list consisted of 20 practice stimuli, followed by two experimental lists consisting of 14 items each. The order of items was random. As in Experiment 1, the participants controlled the amount of time that each item appeared (by pressing the space bar to proceed), but they were told not to spend too much time studying any one item. During the study phase, words were presented one at a time in lowercase letters in the center of the computer screen. At the end of each list, the participants received a "Recall Now" message. At that point, they wrote their responses on a sheet of paper provided by the experimenter.

Design. In each experiment, two factors were crossed in a mixed design. Stimulus type (Type 1, Type 2) was manipulated within

subjects and list type (pure, mixed) was manipulated between subjects. Proportion recalled served as the dependent measure.

Results and Discussion

The mean proportion recalled is presented in Table 3. ANOVA by subjects was performed separately on proportion recalled. If a significant effect was observed in the subjects analysis, a subsequent items analysis was conducted in order to ensure that the effect was consistent across items. In the analysis of items, orthographic neighborhood size was included as a covariate. All effects are reported to be significant at $p < .05$, unless otherwise noted.

As in Experiment 1, there was a clear effect of the number of PO friends in Experiment 2A; the main effect of stimulus type was significant by both subjects [$F_s(1,46) = 9.33, MS_e = .01$] and items [$F_i(1,25) = 9.14, MS_e = .02$]. List type was not significant (both $F_s < 1$). In Experiments 2B and 2C, there were no significant effects (all $p_s > .29$ and all $F_s < 1$, respectively).

Despite the increased emphasis on retrieval operations in free recall versus recognition, the results of the second set of experiments are remarkably similar to those of Experiment 1. Again it is clear that the distinctive property that plays an important role in memory performance is the number of a word's PO neighbors. Specifically, low-N words were recalled at a higher rate than were high-N words. Furthermore, no additional influence was observed for phonological-to-orthographic or orthographic-to-phonological consistency.

EXPERIMENT 3 Recall Memory

One additional experiment was conducted to ensure that the results of Experiment 2A were due to PO neighborhood size and not to encoding time. This experiment was deemed necessary because we failed to record the amount of time each item was studied in Experiment 2A. Therefore, it was possible that better recall performance for low-N words was due to increased study time for these words. In Experiment 3, each word appeared for 2 sec during study.

Table 3
Results From Experiments 2A, 2B, and 2C

	Recalled	
	Proportion	SE
Experiment 2A		
OPC and POC–Low-N	.49	.03
OPC and POC–High-N	.42	.02
Experiment 2B		
OPC and POC–Low-N	.42	.03
OPC and POI–Low-N	.44	.02
Experiment 2C		
OPC and POI–Low-N	.42	.03
OPI and POI–Low-N	.43	.03

Method

Participants. Forty College of Charleston introductory psychology students participated in Experiment 3 for course credit.

Materials. The words from Experiment 2A were used as stimuli, and the same method for list construction and counterbalancing was employed. No practice list was included in Experiment 3.

Procedure. The same procedure was used as in Experiment 2A, with two exceptions. First, each stimulus appeared for 2 sec. Second, the words were presented in black on a white background. E Prime (Psychology Software Tools) was used to control stimulus presentation.

Design. The design was the same as in Experiment 2A.

Results and Discussion

The mean proportion recalled is presented in Table 4. The same analyses and significance levels were used as in Experiment 2A.

As in Experiment 2A, there was a clear effect of the number of PO friends in Experiment 3, the main effect of stimulus type was significant by both subjects [$F_s(1,38) = 5.90, MS_e = .02$] and items [$F_i(1,25) = 5.09, MS_e = .02$]. List type was not significant (both $F_s > 1$).

GENERAL DISCUSSION

Our results indicate that the number of PO neighbors of a word influences both recognition and recall. In Experiments 1 and 2, three conditions of words had a small number of PO neighbors, and performance among these three types was indistinguishable and clearly better than the one in condition of words that had a large number of neighbors. Interestingly, neither PO consistency nor OP consistency produced an additional benefit. In Experiment 3, we replicated the PO neighborhood size effect in recall under an experimenter-paced presentation rate.

The interference framework we outlined in the introduction predicts the results. Specifically, words that share PO neighborhoods with a large number of words experience interference due to the coactivation of phonological and orthographic representations. This interference hypothesis is consistent with the operation of current models of word recognition. Specifically, in both localist and distributed models, words that share similar representations are expected to interfere with one another. Furthermore, the more words that overlap with a target word, the more interference that can be expected.

One might expect that in addition to higher hit rates for words with fewer PO neighbors, false alarm rates would be lower. We do note that if the false alarms for the three conditions of words are combined with few PO neighbors, there is a trend in the predicted direction (.18 to .22). Perhaps the null effect in false alarms is related to how lexical and episodic memories interact. Specifically, the PO neighborhood size effect occurs for hits because the memory trace originates from activation in lexical memory, and the trace is weaker for words sharing representations with a large number of PO neighbors. For false alarms, a memory trace is not accessed, and the activation level alone may not be enough to drive the effect.

Table 4
Results from Experiment 3

	Recalled	
	Proportion	SE
OPC and POC–Low-N	.38	.02
OPC and POC–High-N	.31	.02

Because the results for recognition memory are qualitatively similar to those for recall, we propose that diffusion of activation during encoding may be responsible for the results. Since recall places more of an emphasis on retrieval processes, one would expect that if the PO neighborhood size effect were due to retrieval blocking, then it would be stronger in recall than in recognition.

It is possible that a relationship exists between PO neighborhood size and orthographic distinctiveness—specifically, that words with fewer PO neighbors may be more orthographically distinct than words with more PO neighbors. We addressed this possibility by obtaining normative data on a 1–9 scale for the words used in the present study across 30 participants (see Zechmeister, 1969). Although we found a small difference between conditions (PO–low-N = 4.65 and PO–high-N = 4.36), this difference was not significant [$t(26) = 1.75, p > .09$]. Furthermore, when the orthographic distinctiveness rating was entered as a covariate in the items analyses, all of the reported comparisons involving the number of PO friends remained significant. Finally, if orthographic distinctiveness were driving the results, then we might expect the effect of PO friends to interact with list type in the recall memory experiment. Hunt and Elliot (1980) found that the effect of orthographic distinctiveness occurred when distinct and indistinct words were presented in the same list but not when they were presented in separate lists. In contrast, we found that the number of PO friends did not interact with list type in either Experiment 2A or Experiment 3. This outcome is consistent with the idea that the interference caused by PO neighborhood size occurs in lexical memory and is not due to list characteristics.

Finally, regarding OP consistency, it is important to note that we failed to obtain the pattern of results reported by Hirshman and Jackson (1997). There are three possible reasons for this discrepancy. First, during encoding, Hirshman and Jackson had participants read their stimuli aloud, whereas our participants read these items silently (and we did not observe any effects of consistency on encoding times). Therefore, there may have been less of an influence of the OP translation process in our methodology than in theirs. In fact, in the word recognition literature, OP consistency effects are stronger in reading aloud tasks than in lexical decision tasks where a naming response is not required (e.g., Cortese, 1998). Second, we explicitly informed our participants that their memory for the to-be-read items would be tested, whereas Hirshman and Jackson did not do so with their participants. Thus, it is possible that OP consistency effects arise when

an incidental memory test is employed. Third, we analyzed the items used by Hirshman and Jackson, and as it turns out, OP consistency was confounded with PO neighborhood size in their stimulus set. Specifically, the inconsistent words in the Hirshman and Jackson study had fewer PO neighbors ($M = 1.3$) than did the consistent words [$M = 11.4, t(24) = 5.90, p < .01$]. Thus, it is possible that Hirshman and Jackson's results were actually due to PO neighborhood size and not OP consistency.

We note that because of the methodological differences between our study and that of Hirshman and Jackson, we cannot unequivocally account for the discrepancy between the results of the two studies. Furthermore, our study was not designed to be a conceptual replication of theirs. Instead, we demonstrated a novel finding with respect to PO neighborhood size across three experiments, with many participants, across recognition and recall. Future research will need to determine whether OP consistency effects arise when different experimental situations are employed.

In conclusion, we found a clear effect of phonological-to-orthographic neighborhood size on both recognition and recall. These results suggest that orthographic and phonological representations are coactivated during encoding and/or retrieval, and that either low-N items are more easily differentiated from high-N items, or high-N items receive interference from similarly spelled neighbors, or both.

REFERENCES

- BERNDT, R. S., REGGIA, J. A., & MITCHUM, C. C. (1987). Empirically derived probabilities for grapheme-to-phoneme correspondences in English. *Behavior Research Methods, Instruments, & Computers*, **19**, 1-9.
- CLARK, H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning & Verbal Behavior*, **12**, 335-339.
- COLTHEART, M., DAVELAAR, E., JONASSON, J. T., & BESNER, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535-555). Hillsdale, NJ: Erlbaum.
- COLTHEART, M., RASTLE, K., PERRY, C., LANGDON, R., & ZIEGLER, J. (2001). A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, **108**, 204-256.
- CORTESE, M. J. (1998). Revisiting serial position effects in reading. *Journal of Memory & Language*, **39**, 652-665.
- CORTESE, M. J., BALOTA, D. A., SERGENT-MARSHALL, S. D., & BUCKNER, R. L. (2003). Sublexical, lexical, and semantic influences in spelling: Exploring the effects of age, Alzheimer's disease and primary semantic impairment. *Neuropsychologia*, **41**, 952-967.
- CORTESE, M. J., & SIMPSON, G. B. (2000). Regularity effects in word naming: What are they? *Memory & Cognition*, **28**, 1269-1276.
- HIRSHMAN, E., & JACKSON, E. (1997). Distinctive perceptual processing memory. *Journal of Memory & Language*, **36**, 2-12.
- HUNT, R. R., & ELLIOTT, J. M. (1980). The role of nonsemantic information in memory: Orthographic distinctiveness effects on retention. *Journal of Experimental Psychology: General*, **109**, 49-74.
- KUČERA, H., & FRANCIS, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- PLAUT, D. E., MCCLELLAND, J. L., SEIDENBERG, M. S., & PATTERSON, K. E. (1996). Understanding normal and impaired reading: Computational principles in quasi-regular domains. *Psychological Review*, **103**, 56-115.

- SCHMIDT, S. R. (1991). Can we have a distinctiveness theory of memory? *Memory & Cognition*, *19*, 523-542.
- WALLACE, W. P. (1965). Review of historical, empirical, and theoretical status of the von Restorff phenomenon. *Psychological Bulletin*, *63*, 410-424.
- ZECHMEISTER, E. B. (1969). Orthographic distinctiveness. *Journal of Verbal Learning & Verbal Behavior*, *8*, 754-761.
- ZIEGLER, J. C., STONE, G. O., & JACOBS, A. M. (1997). What is the pronunciation for *-ough* and the spelling for /u/? A database for computing feedforward and feedback consistency in English. *Behavior Research Methods, Instruments, & Computers*, *29*, 600-618.

NOTES

1. *Bulb* was used as an OPC and POC-low-N word in Experiments 1 and 3, and *bulk* was used in Experiments 2A and 2B. The frequency of OPC and POC-low-N words is 6.2 when *bulk* is used. All other values remain the same.

2. *Hurl* and *pearl*, as well as *womb* and *plume*, contained the same phonological rime, but neither pair contained the same orthographic rime. With these exceptions, no two words contained the same phonological rime, and no two words (including filler items) contained the same orthographic rime.

3. An analysis of the spelling errors produced during recall shows that the results were not affected by the spelling accuracy of the responses in any of the experiments reported. In Experiment 2A, slightly more spelling errors were made for low-N words than for high-N words (.01 vs. .00). In Experiment 2B, slightly more errors were produced for OPC and POC-low-N words than for OPC and POI-low-N words (.02 vs. .01). In Experiment 2C, more errors were produced for OPC and POI-low-N words than for OPI and POI-low-N words (.06 vs. .02). We note that by adding the proportion of spelling errors to the means of each condition, the pattern of results remains the same, and the overall level of recall is fairly constant across experiments (Experiment 2A = .46, Experiment 2B = .45, and Experiment 2C = .47). In Experiment 3, spelling errors were very uncommon for either type of stimulus (.00 vs. .00).

APPENDIX

Words Used in the Experiments	Hits	FAs	Hits - FAs	Recall Rates			
				Exp. 2A	Exp. 2B	Exp. 2C	Exp. 3
OPC and POC-Low-N							
pulp	.88	.53	.35	.46	.27		.38
branch	.59	.00	.59	.46	.46		.22
scalp	.71	.12	.59	.46	.33		.20
mount	.82	.59	.23	.53	.46		.50
welsh	.88	.29	.59	.40	.40		.30
garb	.88	.24	.64	.48	.35		.28
filth	.88	.06	.82	.48	.35		.25
babe	.76	.12	.64	.71	.63		.55
elk	.82	.06	.76	.61	.56		.50
leash	.88	.06	.82	.44	.48		.38
void	.82	.12	.70	.40	.38		.43
bulb	.82	.24	.58				.38
bulk				.42	.48		
belch	1.0	.41	.59	.48	.42		.40
cuspl	.82	0	.82	.34	.31		.30
OPC and POC-High-N							
mound	.88	.41	.47		.54		.55
slick	.47	.24	.23	.25			.30
split	.59	.18	.41	.40			.22
scrub	.94	.35	.59	.40			.30
cab	.65	.06	.59	.59			.50
tuck	.71	.00	.71	.38			.22
swell	.53	.12	.41	.52			.32
prong	.76	.24	.52	.29			.28
flog	.82	.12	.70	.57			.25
crust	.35	.29	.06	.29			.23
hunch	.71	.29	.42	.32			.23
trump	.82	.41	.41	.21			.18
lag	.76	.35	.41	.40			.22
fang	.71	.00	.71	.40			.32
OPC and POI-Low-N							
plume	.88	.29	.59		.29	.44	
nil	.82	.00	.82		.50	.46	
odd	.47	.12	.35		.60	.48	
shriek	.88	.12	.76		.48	.35	
roar	.47	.00	.47		.46	.40	
pert	.88	.41	.47		.29	.35	
newt	.94	.24	.70		.67	.58	
dirge	.82	.06	.76		.25	.17	
hype	.94	.18	.76		.33	.31	
cheese	.82	.06	.76		.73	.71	

APPENDIX (Continued)

Words Used in the Experiments	Hits	FAs	Hits – FAs	Recall Rates			
				Exp. 2A	Exp. 2B	Exp. 2C	Exp. 3
OPC and POI–Low-N (Continued)							
pyre	1.0	.12	.88		.33	.44	
hurl	.88	.29	.59		.54	.52	
zinc	.94	.00	.94		.38	.44	
lewd	.71	.24	.47		.31	.25	
OPI and POI–Low-N							
plaid	.65	.24	.41			.35	
yacht	.88	.06	.82			.50	
ache	.82	.12	.70			.52	
ski	.82	.00	.82			.60	
niche	.82	.41	.41			.33	
bead	.88	.24	.64			.52	
sieve	1.0	.00	1.0			.33	
pearl	.88	.47	.41			.56	
crepe	.65	.06	.59			.27	
gauge	.94	.65	.29			.29	
ghoul	.94	.00	.94			.31	
womb	.88	.06	.82			.50	
monk	.82	.12	.70			.52	
tongue	.76	.18	.58			.46	

(Manuscript received May 27, 2003;
revision accepted for publication December 5, 2003.)