Fractal Characteristics of Response Time Variability

John G. Holden

Department of Psychology
California State University, Northridge

The relations between English spellings and pronunciations have been described as a fractal pattern. Manipulations of word properties are constructed to coincide with the fractal pattern of ambiguity in these relations (sampled as random variables). New word naming and lexical decision experiments replicate previously established effects of relations between word spellings and pronunciations. The structure of these relations establishes a measurement scale, which is nested in a manner that is loosely analogous to the way that centimeters are nested within meters, and so on. The experiments revealed that broader, increasingly stretched—more variable—distributions of pronunciation and response time accompany the introduction of higher resolution measurement scales to the naming and lexical decision measurement process. Moreover, the distribution of lexical decision response times is shown to obey an inverse power-law scaling relation, which implies lexical decision does not conform to a characteristic measurement scale.

Fractal patterns in nature are composed of nested forms, which cannot be measured on a single scale of measurement. The result of a measurement will depend on the size of the increment, or scale, used in the process of measurement (Bassingthwaighte, Liebovitch, & West, 1994; Mandelbrot, 1983; Schroeder, 1991). For example, the measured length of the British coastline will increase proportionally if the "yardstick" is shortened from kilometers to meters. An even shorter centimeter measurement scale would result in a further proportional increase in the measured length of the coastline. The changing measurements result from using regular line segments, the yardstick, to approximate the nested *self-similar* structure of the coastal bays and peninsulas. The length measurement increases when a

---

bay or peninsula that is not captured at a lower resolution adds length at a higher resolution. Thus, as smaller and smaller sub-bays and subpeninsulas are resolved, they add to the length of the coastline.

When measurements change as a function of measurement scale, there is no "true" or characteristic value for the measurement. The length of the British coastline grows as a power function of the precision of the yardstick used to measure length (plus a constant). This power-law scaling relation implies that results of a measurement depend on the measurement scale or sampling unit used to take the measurement (over a finite range of scales). *Power-law scaling relations,* linear relations between the logarithms of the scale and the logarithms of the measurement result, are commonly observed of natural phenomena described using fractal geometry and are symptomatic of self-similar patterns (Bassingthwaighte et al., 1994). In mathematical fractals, this nonlinear, multiplicative measurement amplification can appear at an infinite range of scales, a precise form of self-similarity. In natural fractals, the most common form of self-similarity appears to be statistical self-similarity. Nevertheless, the result of a measurement of a natural fractal is also amplified in proportion to the measurement scale.

Two naming experiments follow that measure variability in pronunciation times, the time to initiate a pronunciation from a word's spelling. The experiments contrast words that have nested sources of ambiguity in the relations between their spellings and pronunciations. Ambiguity is conceived here as abstractly parallel to the nested structure of a coastline. A coastline is jagged at several scales: Large bays nest smaller bays within them, and so on. Likewise, the pronunciation of words' spellings are ambiguous at several scales, as I explain in the next section. Just as length accumulates as a scale gets access to successively nested bays, variability in pronunciation time is expected to grow as ambiguity manipulations access the nested ambiguity in the relation between words' spellings and pronunciations.

Similarly, two lexical decision experiments measure the time to recognize a word as a word—the time to respond that a spelling pattern is a legitimate English word. In this case, ambiguity appears at different resolutions in the relations between words' pronunciations and their spellings—now emphasizing the inverse mapping from pronunciation to spelling. As in naming, the measure of variability in recognition time is expected to grow as nested sources of ambiguity accrue in the relations between pronunciation and spelling. Response-time distributions that are increasingly stretched by nested ambiguity may be expected if word naming and recognition performance are fractal processes. In contrast to the fractal perspective, conventional word-recognition theories offer no a priori basis for systematic changes in response-time variability.

Standard inferential statistical procedures assume that, in the long run, measured (dependent) variables yield approximately symmetrical Gaussian distributions. As a consequence, the term *variability* is often used to refer specifically to the variance or standard deviation of a distribution in that framework. I use the term variability more inclusively—more variable distributions encompass a wider range

of numerical values. Thus, increases in variability loosely refer to simultaneous increases in parameters, such as variance and skew. Ultimately, however, increasing variability refers here to a pattern of progressive "stretching" in the shapes of density functions of pronunciation and lexical decision times.

SPELLING AND PRONUNCIATION

Van Orden, Pennington, and Stone (2001) described the relations between English spellings and pronunciations as a fractal pattern. They described relations at three scales, grapheme-phoneme, body-rime, and whole-word, which I illustrate next.

Ambiguous Graphemes

The grapheme *B,* at the beginning of a word, is always pronounced /b/, and the phoneme /b/ is always spelled *B.* This defines an invariant, symmetrical relation between a grapheme and a phoneme. Some consonant graphemes and phonemes have invariant relations, but vowel spellings are always ambiguous. Ambiguity implies a spelling that can be pronounced in more than one way (or phonemes that can be spelled in more than one way). Vowel spellings can always be pronounced in multiple ways. For example, *duck, burn,* and *dude* include multiple pronunciations of the spelling _ u _ _. Similarly, the phoneme /u/ can be spelled _ u _ _, _ o _ _, or _ o e _ , respectively, in *duck, monk,* and *does.* These examples illustrate the two forms of ambiguity in vowel graphemes and phonemes. Ambiguous vowel graphemes map onto more than one phoneme, and ambiguous vowel phonemes map onto more than one grapheme.

Ambiguous Spelling Bodies

Ambiguous vowel graphemes sometimes rely on contextual constraints to specify a single pronunciation option. An example is the context provided for the spelling _ u _ _ by the relation between a word's spelling body (e.g., _uck) and its pronunciation rime (e.g., /uk/). The word *duck* includes an invariant, symmetrical relation between its spelling body, _uck, and its pronunciation rime, /_uk/. All words that contain the spelling body _uck are pronounced to rhyme with *duck* (e.g., *luck* and *buck*), and all words that are pronounced to rhyme with /_uk/ are spelled using the same _uck spelling pattern.

The vowel letter _ u _ _ entails a large set of potential relations between spelling and pronunciation, which subdivides into smaller sets of relations, one of which corresponds to _uck ↔ /uk/. It is in this sense that a body-rime relation is nested within a large set of potential vowel relations. Invariant body-rime relations such as _uck ↔ /uk/ collapse the ambiguity of the vowel relation into a single rime pronun-

ciation. Invariant body-rime relations limit ambiguity to the scale of grapheme-phoneme relations. Body-rime invariant words only require that grapheme ambiguity be resolved to specify a single pronunciation option. Thus, pronunciation time variability based on these words reflects ambiguity at the grapheme scale.

Invariant body-rime relations are relatively rare in English because most spelling bodies are ambiguous (Stone, Vanhoy, & Van Orden, 1997). The spelling body _int in *pint,* for example, is pronounced /aɪnt/ in *pint* but also maps to /ɪnt/ as in *mint.* An ambiguous spelling body reveals ambiguity over and above vowel ambiguity at the grapheme scale. Body-rime ambiguous words require that grapheme ambiguity and spelling-body ambiguity be resolved to specify a single pronunciation option. Pronunciation time variability based on these words reflects ambiguity at the grapheme scale and the spelling-body scale.

Ambiguous Whole Words

Ambiguity in the _int vowel pronunciation is resolved once the context supplied by the whole-word spelling *p-i-n-t* is taken into account. The whole word *pint* has an invariant, symmetrical relation between its whole-word spelling and its whole-word pronunciation. *Pint* has only one correct pronunciation, and its pronunciation only has one legitimate spelling. This invariant symmetrical relation is "contained" within the set of potential body-rime relations for _int. Taking into account the whole-word spelling has further subdivided _ i _ _ and _int's potential relations into a particular relation, *pint* ↔ /paɪnt/. Invariant whole-word relations limit ambiguity to the grapheme and spelling-body scales.

Most English words have invariant whole-word relations, but there are important exceptions. Homographs are whole words with more than one pronunciation. For example, the homograph *lead* has two legitimate pronunciations: one rhymes with *bead,* the other with *head.* This illustrates ambiguity in a whole-word spelling. In a laboratory setting that presents *lead* in relative isolation, a unique vowel pronunciation cannot be resolved from available episodic constraints—they would be balanced between the two pronunciations. Historical sources of constraint may tip the balance, however. A participant's unique history of literate discourse, and consequent internalized constraints, will favor one or the other pronunciation (cf. Kawamoto & Zemblidge, 1992). Homographs reveal ambiguity at all three scales of the relations between spelling and pronunciation. Thus, pronunciation time variability based on these words reflects ambiguity at the grapheme scale, the spelling-body scale, and the whole-word spelling scale.

The overall pattern of relations between English spelling and pronunciation is proposed to be a fractal pattern of ambiguity. Ambiguous grapheme-phoneme relations can entail nested, ambiguous body-rime relations that, themselves, can entail nested ambiguous whole-word relations. As noted, homographs, such as *lead,* are ambiguous at all three scales of the nested relations between spelling and pronunciation.

SPELLING, PRONUNCIATION,
AND THE BRITISH COASTLINE

With respect to the coastline analogy, a word in which ambiguity is limited to the grapheme scale would correspond to a relatively low-resolution "kilometer scale" to measure how vowel ambiguity affects pronunciation time variability. In contrast, spelling-body scale vowel ambiguity amplifies grapheme scale ambiguity and pronunciation time variability, compared to a word with a spelling body that specifies a single pronunciation option. It reveals more of the process that resolves ambiguity in performance measures. Similarly, homographs amplify vowel ambiguity and pronunciation time variability, over and above ambiguity at the spelling-body scale, and reveal still more of the process that ends in a pronunciation.

The empirical goal of this research is to track changes in pronunciation and lexical decision time variability that accompany the separate scales of relations between word spellings and pronunciations. Distinguishing the perspective of a *means analysis* from the perspective of a *variability analysis* serves as an entry point for the hypotheses that follow. Pronunciation times and lexical decision times are typically viewed as random variables. Presenting particular words under particular circumstances yields a range or distribution of response-time measurements. A means analysis assumes the existence of a characteristic pronunciation time that is shrouded by unsystematic, additive sources of noise. By contrast, a variability analysis is concerned with identifying systematic changes in variability; it assumes that the pattern of changes in variability is informative, that it may reflect the intrinsic dynamics of the system (cf. Riley & Turvey, in press). Just as I might ask, "What is the length of the British coastline?" and get different answers with different scales—so may I ask, "How variable are response times?" and get different answers from different scales of ambiguity.

My main prediction is that using progressively higher resolution scales to measure performance will yield a pattern of progressive increases in the parameters that describe variability—that is, the shape of pronunciation and response-time distributions. Changing scales will not simply shift a distribution's mean. Visually, this will appear as a progressive "stretching" of a response-time distribution toward very slow response times. Each scale of resolution in ambiguous relations, sampled as random variables, should amplify the variability observed using a lower resolution ambiguity scale. A similar pattern can be observed in the distributions of *synthetic* pronunciation times generated by recurrent (iterative) network models of word naming (I. Choi, personal communication, October 31, 2001). The reason that these patterns emerge in a model's behavior motivate my predictions concerning human performance.

Multiple, mutually inconsistent relations (connections) among nodes are an iterative model's analogue to vowel ambiguity. Iterative networks multiply a vector of node values by a matrix of connection weights, and the result serves as the next input vector. Past that point, it continues to iterate until a relatively consistent, sta-

ble pattern of node values emerges. This means that iterative models rely on multiplicative interactions to stabilize their dynamics. As such, an iterative model's distribution of finishing times reflects multiplicative interactions among random variables. The product of two or more random variables yields a lognormal distribution, a point I return to later (Ulrich & Miller, 1993; West & Deering, 1995).

All else being equal, active, mutually inconsistent node relations decrease the probability that a model will stabilize its interaction on a given iteration. Over many synthetic trials, competing relations may reveal a wider range of a model's finishing times because reducing the probability that a model will stabilize on any particular iteration does not preclude it from doing so in the long run. This is one basis of my prediction that access to multiple scales of mutually inconsistent pronunciation options, for example, may amplify variability in empirical distributions of response times.

EXPERIMENT 1: GRAPHEME SCALE AND SPELLING-BODY SCALE AMBIGUITY

Experiment 1 uses a word naming task, which measures the time required to pronounce individual words and replicates the standard effect of spelling-body ambiguity (*pint,* hereafter called the spelling-body scale), in a contrast with spelling-body invariant words (*duck,* hereafter called the grapheme scale; Glushko, 1979; Jared, McRae, & Seidenberg, 1990). It tests whether the conventional mean effect of spelling-body scale ambiguity is better described as a skewing, or more precisely, a stretching of pronunciation time distributions toward slower times—an increase in variability. This hypothesis is tested by contrasting distributions of pronunciation times at the grapheme scale (*duck*) with those at the spelling-body scale (*pint*). This experiment was constructed to closely mimic previous demonstrations of this effect.

In this, and all following experiments, contrasts were accomplished by item yoking. Each grapheme scale item was yoked to a spelling-body scale item. Yoked item pairs allowed for within-subject distribution analyses; each participant served as his or her own control. Yoked items were matched with respect to variables, other than spelling-body ambiguity, that affect pronunciation time. Only item pairs that both produced correct pronunciation times, from the same participant, were included in statistical analyses. (In every experiment, analyses that admitted error pronunciations and response times yielded the same pattern of results.)

Method

Participants.   Thirty-five introductory psychology students, who were all native English speakers, participated in the experiment.

Materials.   The spelling-body scale of ambiguity is defined with respect to word bodies (Jared et al., 1990). The key stimuli consisted of 21 yoked pairs of grapheme scale (*duck*) and spelling-body scale (*pint*) items. All items were monosyllabic four- and five-letter English words.

A word frequency count estimates the rate of recurrence (per million words) of individual words in large samples of printed text. Words that have high frequency counts appear more often in samples of printed text than do words with low frequency counts. Frequently encountered words are responded to more quickly and accurately in a variety of word-recognition tasks, such as word naming and lexical decision. I yoked the grapheme scale and spelling-body scale words with respect to their frequency counts: Grapheme Scale, $M = 4.95$ ($SD = 4.19$); Spelling-Body Scale, $M = 5.33$ ($SD = 4.78$). I also yoked items with respect to their number of letters: Grapheme Scale, $M = 4.38$ ($SD = 0.50$); Spelling-Body Scale, $M = 4.43$ ($SD = 0.51$). In addition, a voice key takes different amounts of time to detect different initial phonemes, so I matched the words on that basis as well.

Word bodies are spelling patterns (e.g., _int in *mint* and *pint*) that define a spelling neighborhood. A *spelling neighborhood* is composed of all monosyllabic words that include the same spelling body. For example, *pint, mint, lint,* and all other words ending in _int comprise _int's spelling neighborhood. Ambiguous spelling bodies, such as _int, map to more than one pronunciation rime. Body-rime relations can be divided into dominant and subordinate relations. For example, with the exception of *pint,* every word in _int's spelling neighborhood uses the _int ↔ /Int/ relation. Thus, _int ↔ /Int/, as in *mint,* is _int's dominant body-rime relation. By contrast, _int ↔ /aɪnt/, as in *pint,* is _int's subordinate body-rime relation.

A *pronunciation ratio* estimates the relative dominance of a body-rime relation. The pronunciation ratio is computed by dividing the sum of the frequencies of words that share a body-rime relation into the overall sum of the frequencies of all words that have the spelling body. In cases where there are exactly two possible pronunciations, and this ratio is less than one half, the body-rime relation is a subordinate relation. In cases where there are several alternative body-rime relations in a spelling neighborhood, the subordinate relation is the relation with the smallest pronunciation ratio.

Words that are both body-rime and *rime-body* invariant are relatively rare in English, which severely restricted the available items (Stone et al., 1997). Rimebody ambiguity refers to a pronunciation rime that has multiple spellings (e.g., /ûrn/, as in *fern,* could be spelled _urn, as in *turn*). Although all items were strictly body-rime invariant, the definition of rime-body *invariance* was relaxed to rimebody *dominance* for 7 grapheme scale items and 12 spelling-body scale items. Thus, rime-body dominant items appeared in similar proportions at both scales. The relaxed definitions were justified, in large part, because rime-body ambiguity does not appear to reliably affect pronunciation time (Ziegler, Montant, & Jacobs, 1997). Seven of the 21 grapheme scale items had alternate rime spellings in words with frequencies of occurrence greater than 1 per million. Most of these words were

relatively obscure, with frequencies of 7 or less per million (Kučera & Francis, 1967). Two notable exceptions were the target *weed,* with rime /ed/, which may also be spelled _ead, as in *bead,* and the target *quake,* with rime /ak/, which may also be spelled as in *break.* Of the 12 spelling-body scale targets, 10 word rimes had alternative spellings in words with frequency counts greater than 10 per million. Again, rime-body status has, so far, failed to reliably affect pronunciation time, so these items are not expected to influence the outcome of the experiment on the basis of their rime-body status.

Almost all the spelling-body scale targets were chosen to have subordinate body-rime relations, *rouge* and *wharf* being the exceptions. Also, with one exception, the target *broth,* I used no items whose whole-word dominant pronunciation ratio was derived from a single very high frequency word. Body-rime invariant items that have large spelling neighborhoods, with several high-frequency members, were preferred in the selection of grapheme scale items. For example, *duck* has several high-frequency neighbors (e.g., *truck, luck,* and *stuck*). Hermit words that are themselves the only member of a spelling neighborhood were not used as targets. None of the target spelling-bodies appeared more than once in the experiment. The yoked pairs appear in Appendix A.

Procedure. A trial began with a fixation stimulus (+++) presented for 415 msec at the center of a standard PC screen. The fixation was followed by a blank screen for 200 msec, after which a target word appeared. Participants named aloud the target word, quickly and accurately (as instructed). A voice key registered the participant's pronunciation time, and the target remained visible until the next trial was initiated. The participant pressed one of two keys to initiate the next trial. A *correct* key signaled a successful trial, an *error* key signaled that they had mispronounced the word. An intertrial interval of 1,200 msec occurred following the key press. (The naming procedure was modeled after Spieler & Balota, 1997.)

During the experiment, the experimenter sat out of the participant's view and transcribed errors. This ensured that an improper "error" key press by the participant did not jeopardize coding. Also, some words, such as *hoof,* are pronounced differently in different regions of the country. Thus, on occasion, a participant produced an ambiguous pronunciation. These pronunciations were verified by having the participants pronounce the item again, the way they typically pronounce it, at the end of the experiment.

Participants began by completing 50 randomly ordered practice trials and received feedback from the experimenter during the practice trials concerning what constitutes an error. This was followed by 5 fixed-order buffer trials and a 140-trial experimental block. The 42 experimental stimuli were embedded in a set of 98 filler items. The items were presented in a pseudorandom order, which ensured targets were randomly but evenly distributed among subsets of fillers. The entire experimental procedure required approximately 20 min to complete.

Results and Discussion

Variability analysis.   I used one-tailed significance tests, with alpha set at .05, for all planned contrasts in this and all experiments that follow, because all involve directional hypotheses. As noted in the introduction, the higher resolution spelling-body scale is expected to amplify the parameters that describe the shapes of the pronunciation time distributions (i.e., variability and skew). Evidence of this pattern is revealed in a stretching analysis. I could not directly test the stretching hypothesis with standard variance and skew statistics, however. Higher moments of response-time distributions are difficult to estimate because they fluctuate wildly (Ratcliff, 1979). This is to be expected from attempts to measure a process that has no characteristic scale of measurement. Furthermore, variance (and skew) are computed relative to a distribution's own mean (and variance), and meaningful contrasts require that both distributions have the same mean and variance.

Instead, I recorded the first and third quartiles of each participant's distribution of correct grapheme scale and spelling-body scale pronunciation times. A pattern of stretching in the spelling-body scale distribution, with respect to the grapheme scale distribution, is established if the difference between the third quartiles of the two distributions is reliably larger than the difference between the same distribution's first quartiles. This pattern would appear as an overadditive interaction in a 2 × 2 repeated measures analysis of variance (ANOVA), which crosses the two ambiguity scales with the two quartile measures. The interaction term of an ANOVA that treated participants as a random variable was statistically reliable, $F_p(1, 34) = 14.30, p < .05$, as was the interaction term for an ANOVA that treated items as a random variable, $F_i(1, 20) = 10.94, p < .05$.

I used planned contrasts to reveal pattern of the interactions. On average, the difference between the first quartiles of the participant analysis grapheme and spelling-body scale distributions was 24 msec and, as predicted, the average difference between the third quartiles of the same distributions was larger, at 62 msec. Both differences were statistically reliable, $t_p(34) = 3.54, p < .05$, and $t_p(34) = 5.53, p < .05$, respectively. Likewise, the average difference of 21 msec between the first quartiles of the item-analysis grapheme and spelling-body scale distribution was not statistically reliable. However the 82 msec difference between the third quartiles of the same two distributions was statistically reliable, $t_i(20) = 3.84, p < .05$. The stretching pattern established by these analyses corroborates the hypothesis that variability in pronunciation time is amplified as ambiguity scales—the spelling-body scale—access nested ambiguity in the relation between word spellings and pronunciations.

Multiplicative interaction.   The stretching pattern is readily visible in the pronunciation time distributions. I estimated probability density functions for the grapheme scale and spelling-body scale distributions by transforming the empirical pronunciation time distributions into smoothed probability density curves with a

nonparametric lognormal kernel-density estimator (Silverman, 1986). A lognormal kernel density represents each pronunciation time measurement as a small lognormal curve, rather than a point, centered about its recorded value. A probability density is estimated by summing (and normalizing) the area occupied by the kernels over successive intervals of pronunciation time. (A lognormal kernel corresponds to applying a symmetric Gaussian kernel to the logarithms of the pronunciation times and then transforming the resulting density back into the linear domain.) The extent of smoothing is controlled by adjusting the standard deviation of the kernel, just as one might adjust the width of a histogram bin. The smoothing parameter was chosen using a formula based on the standard deviation of the pronunciation time distributions (cf. Silverman, 1986; Van Zandt, 2000). I used a lognormal kernel because it progressively increases the amount of smoothing as it moves into the sparse, slow tail of the distribution. This is appropriate for positively skewed distributions. It preserves detail in the dense front end of the distribution while simultaneously compensating for the lower proportion of raw data points in the slow tail of the distribution.

A lognormal empirical density function will closely resemble a standard Gaussian density function after a logarithmic transformation of the raw data. The approximate lognormal shape of response-time distributions has been previously noted (e.g., Ratcliff & Murdock, 1976). The dashed line in Panel A of Figure 1 depicts an ideal Gaussian distribution with equivalent variability, centered about the median
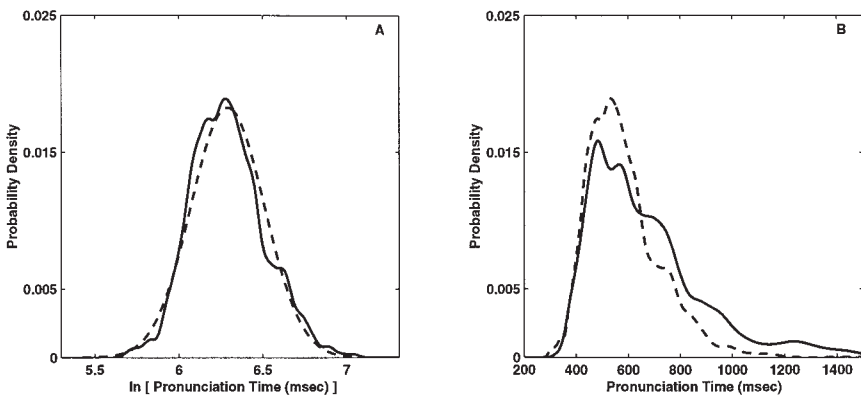

FIGURE 1   The solid line of Panel A pertains to the kernel-smoothed density function of pronunciation times to the grapheme scale items after a logarithmic transformation. The dashed line of Panel A represents an ideal Gaussian curve, with the same standard deviation, centered on the empirical distributions' median. The empirical distributions' approximate Gaussian shape, in the logarithmic domain, indicates a close resemblance to a lognormal distribution and implies multiplicative interaction. Panel B contrasts variability on the grapheme and spelling-body scales, now in the linear domain. The dashed line depicts the variability to grapheme scale words. The solid line depicts the variability at the spelling-body scale. The distribution at the spelling-body scale displays a longer, more stretched tail than that of its grapheme scale counterpart.

of the empirical grapheme scale distribution depicted by the solid line (treating the *x* axis as a linear scale). The close similarity of the grapheme scale distribution to an ideal Gaussian distribution, in the logarithmic domain, could imply that word naming performance entails multiplicative interactions among system parameters, a point I elaborate in the Summary.

The variability changes that accompanied the change in measurement scale are apparent in a visual contrast to the shapes of the two density functions of pronunciation time. Individual parametric estimates of variability, such as variance and skew, describe specific aspects of the shape of a distribution—they are abstracted, shorthand shape descriptors. A visual examination of the entire distribution supplies a more complete account of the shape of a distribution, however.

The change from the grapheme scale to the spelling-body scale amplified the measured variability in the observed probability density functions. Probability density functions for the two scales are plotted in Panel B of Figure 1. The dashed line portrays the distribution of observed variability at the grapheme scale. The solid line portrays the variability on the spelling-body scale. Variability in the pronunciation times at the spelling-body scale appears as a longer, more stretched slow tail compared to the grapheme scale. Although the modes of the two distributions are similar, the spelling-body scale yields a distribution with many more extreme, slow-pronunciation times. Next, I describe a systematic replication, expanded to include contrasts among all three scales of spelling ambiguity.

EXPERIMENT 2: GRAPHEME VERSUS SPELLING-BODY VERSUS WHOLE-WORD SPELLING AMBIGUITY

Experiment 2 estimates variability in pronunciation times on all three scales of vowel ambiguity. Pronunciation time distributions on a whole-word scale, which uses homographs such as *lead* as targets, are contrasted with distributions on a spelling-body scale, which uses words such as *pint* as targets. Pronunciation times on a spelling-body scale are in turn contrasted with pronunciation times on a grapheme scale, using words such as *duck* (a direct replication of Experiment 1). Items were yoked across all three scales, which allows a simultaneous contrast of the variability on the three scales. Previous research has established that homographs yield exaggerated pronunciation times compared to nonhomographic words (Kawamoto & Zemblidge, 1992).

I conducted Experiment 1 to establish a stronger variability contrast between the grapheme and spelling-body scales than I could accomplish in this three-scale contrast. Most homographs are high-frequency words, and performance tends to be asymptotic for high-frequency words (cf. Jared, 1997). This weakens the contrast of the three scales. Furthermore, there are very few homographs in English (compared to Hebrew, for example) and only seven homographs were suitable for

use in this experiment. Consequently, the three-scale contrast relies on fewer items than the two-scale contrast of Experiment 1.

Given the clear demonstration of variability changes at the grapheme and spelling-body scales established by Experiment 1, I gave priority to the control factors between the whole-word spelling scale (homographs) and the spelling-body scale in Experiment 2. Homographs tend to be pronounced with their statistically dominant pronunciation (see pronunciation ratio in Method), which requires a contrast to the dominant pronunciation on a spelling-body scale. More often than not, *mint* would be a proper contrast for most pronunciations of a homograph, but not *pint,* for example. As such, the contrast between variability on the grapheme scale and variability on the spelling-body scale is weaker in this experiment than it was for Experiment 1. Because I could not know in advance which of the alternate pronunciations a participant might provide to a homograph, I selected two different control items for each homograph.

If naming resembles a fractal or iterative process, then admitting more nested scales of vowel ambiguity into the measurement procedure should change the result of the measurement. Each scale of resolution of variability should nonlinearly amplify the variability measurements taken at lower resolutions. The result of the measurement would thus depend on the measurement scale used to measure pronunciation time variability.

Method

Participants.   Thirty-five additional introductory psychology students, who were all native English speakers, participated in the experiment.

Materials.   The key stimuli were 7 four- and five-letter monosyllabic homographs. The critical control variable concerns the relative frequency of the two alternative body-rime relations that may be used to determine a pronunciation for a homograph.

Each homograph has two legitimate pronunciations: a statistically dominant pronunciation, defined as the most frequently occurring body-rime relation and a subordinate body-rime relation corresponding to the less frequently occurring, alternative body-rime relation. For example, the homograph *bass* has a frequency count of 16 occurrences per million. Its statistically dominant pronunciation rhymes with *mass.* Adding the frequencies of all the words that rhyme with *mass,* and that use either the _ass ↔ /as/ or the _ass ↔ /ace/ body-rime relation, including *bass,* yields a total summed frequency count of 603 per million (e.g., *class, pass,* etc.). The statistically subordinate pronunciation of *bass* rhymes with *space,* and it is the only word in the _ass spelling neighborhood that uses the _ass ↔ /ace/ body-rime relation. Its maximum estimated frequency can be no greater than 16 occurrences per million. For each homograph, the subordinate pronunciation ratio was computed by dividing the summed frequency of words that have the subordi-

nate body-rime into the total summed frequency of the words that contain the spelling body. The relative dominance of a body-rime relation was estimated by subtracting the subordinate ratio from one.

If no other words in the homograph's spelling neighborhood used the alternative body-rime relation, the frequency of the homographic letter string itself was used to estimate the subordinate pronunciation frequency. Thus, in the case of *bass* (pronounced to rhyme with *space*), the subordinate body-rime ratio was .03 (i.e., 16/603), and the dominant body-rime ratio was .97 (i.e., 1 − .03). The same procedure was applied to *wound* and *wind;* both letter strings represented the only alternative body-rime relation in their respective spelling neighborhoods.

The controls chosen to yoke with a homograph's dominant pronunciation were spelling-body scale ambiguous words matched, as closely as possible, to the homograph's dominant pronunciation ratio. Likewise, controls for the subordinate pronunciation were matched to the homograph's subordinate pronunciation ratio. The pronunciation that was actually provided for each homograph by each participant determined which control item was used in subsequent analyses.

The dominance assignments for three items were equivocal, because reasonable alternative assignment procedures could have been adopted. The spelling neighborhood for *lead* includes another homograph, *read;* I assigned the frequency of *lead* to one body-rime relation and *read* to the other. The spelling neighborhoods of *dove* and *close* allow three possible body-rime relations, and in the case of *dove* it could be argued that its neighborhood does not have a dominant body-rime relation. In all three cases, the other yoking constraints limited the list of candidate controls to the point that no better matches were available, using either method for computing the dominance ratio. The mean relative dominance ratio was .10 (SD = .07) for the homograph's subordinate pronunciations and .13 (SD = .08) for the subordinate spelling-body scale yokes. The mean relative dominance ratio was .90 (SD = .07) for the homograph's dominant pronunciations and .73 (SD = .10) for the dominant spelling-body scale yokes.

In addition, all the items were matched for length: Grapheme Scale, M = 4.00 (SD = 0.00); Dominant Spelling-Body Scale yokes, M = 4.14 (SD = 0.38); Subordinate Spelling-Body Scale yokes, M = 4.14 (SD = 0.38); Whole-Word Spelling Scale, M = 4.29 (SD = 0.49); and word frequency: Grapheme Scale, M = 57.14 (SD = 68.00); Dominant Spelling-Body Scale yokes, M = 91.14 (SD = 139.63); Subordinate Spelling-Body Scale yokes, M = 57.29 (SD = 79.97); Whole-Word Spelling Scale, M = 69.29 (SD = 84.60). Care was taken to ensure close frequency matches for items with relatively low frequency counts. A minority of words occur often enough to be classified as high-frequency words. In order to increase the pool of available yoking candidates, the frequency-matching criteria for high-frequency items was relaxed. Likewise, the strict initial phoneme-matching criteria were relaxed to allow yoking of categorically similar initial phonemes, which will trigger a voice key in a similar way (S. Goldinger, personal communication, April 6, 2000). The yoked quartets appear in Appendix B.

Procedure.   The timing of target presentations and the practice procedures were identical to Experiment 1. Practice trials were followed by a 134-trial experimental block. It began with 6 fixed-order buffer trials, followed by 126 pseudorandomly ordered trials. Each set of four yoked targets was presented randomly within blocks of fillers, and the blocks themselves were presented in a random order. During the experiment, the experimenter sat out of the participant's view and transcribed homograph pronunciations and errors. The entire experimental procedure required approximately 20 min to complete.

Results and Discussion

Variability analysis.   A pattern of stretching at the spelling-body scale, as compared to the grapheme scale, is established if the difference between the spelling-body scale distribution's third quartiles is reliably larger than the average difference between the same distribution's first quartiles. This would appear as an overadditive interaction in an ANOVA that crosses the two scales with the pronunciation time measures taken at the two quartiles. I tested the participant- and item-interaction terms using 2 × 2 repeated measures ANOVAs. The interaction term for the participant analysis was statistically reliable, $F_p(1, 34) = 4.55, p < .05$; the interaction term for the item analysis failed to reach statistical significance. The null outcome of the item analysis is not surprising. The effects of most standard factors that affect naming are attenuated to high-frequency words (Van Orden, Pennington, & Stone, 1990). Also, the homograph yoking constraints imposed a weaker contrast between these two scales, and after all, there were only seven items at each scale.

Planned contrasts were used to reveal the pattern of the participant-analysis interaction. The average difference of 2 msec between the first quartiles of the grapheme and spelling-body scales was not statistically reliable, but the average difference of 24 msec between the third quartiles of the same distributions was reliable, $t_p(34) = 2.34, p < .05$. I also used planned contrasts to determine whether the statistically unreliable interaction of the item analysis at the grapheme and spelling-body scales may yield an apparent stretching pattern. The average difference between first quartiles of the grapheme scale and spelling-body scale distributions was only 14 msec, although the average difference between the third quartiles of the same distributions was 25 msec. Although the item-analysis contrasts were not statistically reliable, the apparent pattern of the differences is consistent with the statistically reliable stretching pattern of the participant analysis.

An important goal of Experiment 2 was to establish variability changes in a contrast between the spelling-body and whole-word scales. A pattern of stretching for the spelling-body and whole-word scales is established if the difference between the distributions' third quartiles is reliably larger than the difference between the same distributions' first quartiles. I tested two interaction terms, one treated participants as a random variable, the other treated items as a random variable. Both

analyses used a 2×2 repeated measures ANOVA. And both interaction terms were statistically reliable, $F_p(1, 34) = 7.32$, $p < .05$, and $F_i(1, 6) = 14.79$, $p < .05$, respectively.

I again used planned contrasts to reveal the pattern of these interactions. The 21-msec average difference between first quartiles of the spelling-body and whole-word scale participant-analysis distributions was reliable, $t_p(34) = 3.42$, $p < .05$, and the third quartiles of the same distributions yielded a larger, statistically reliable, average difference of 108 msec, $t_p(34) = 3.38$, $p < .05$. Likewise, the average difference of 16 msec between the first quartiles of the spelling-body and whole-word pronunciation scale item distributions was statistically reliable, $t_i(6) = 2.16$, $p < .05$, and the average difference of 46 msec between the third quartiles of the same distributions was larger and also statistically reliable, $t_i(6) = 3.76$, $p < .05$. The result of the variability measurement appears to depend on the scale used to measure variability.

The variability changes that track the three measurement scales are apparent in the probability density functions of the pronunciation times. Estimated probability density functions of the three distributions were obtained using the kernel-density estimation procedure described in Experiment 1. The empirical probability density functions of the three distributions are plotted in Figure 2. The dotted line depicts the pronunciation time variability at a grapheme scale. The dashed line depicts the pronunciation time variability on the spelling-body scale, and the solid line depicts the pronunciation time variability at the whole-word spelling scale. The qualitative pattern revealed by this contrast is identical to that revealed with the variability contrast at the grapheme scale and spelling-body scales in Experiment 1. The distribution of pronunciation times to words at the whole-word spelling scale has a longer, more stretched tail than the distribution to words at the spelling-body scale, which is in turn more stretched than the distribution to words at the grapheme scale. As before, the distributions illustrate a progressive divergence, primarily in their slow tails, to include more and more extreme, slower pronunciation times—a change in variability that is statistically reliable by participants.

This stimulus set was comprised of mostly high-frequency words. High-frequency words are, by definition, more commonly encountered while reading. Thus, in contrast to the low-frequency items of Experiment 1, the high-frequency words' pronunciation times are compressed more toward the fast end of the distribution. This can be seen in the more pointed (leptokurtic) peaks of the distributions in this experiment, compared to the distributions appearing in Experiment 1. Nevertheless, the slower tail of the whole-word scale distribution is stretched beyond that of the two other distributions and contains pronunciation times, even in the interval between 1,000 and 1,500 msec.

To take a measurement of something is to make contact with it in some way. To measure the duration of the act of naming a printed word, I must constrain the act by presenting individual words. Control parameters are sampled probabilistically as particular words are presented to particular participants at particular points in
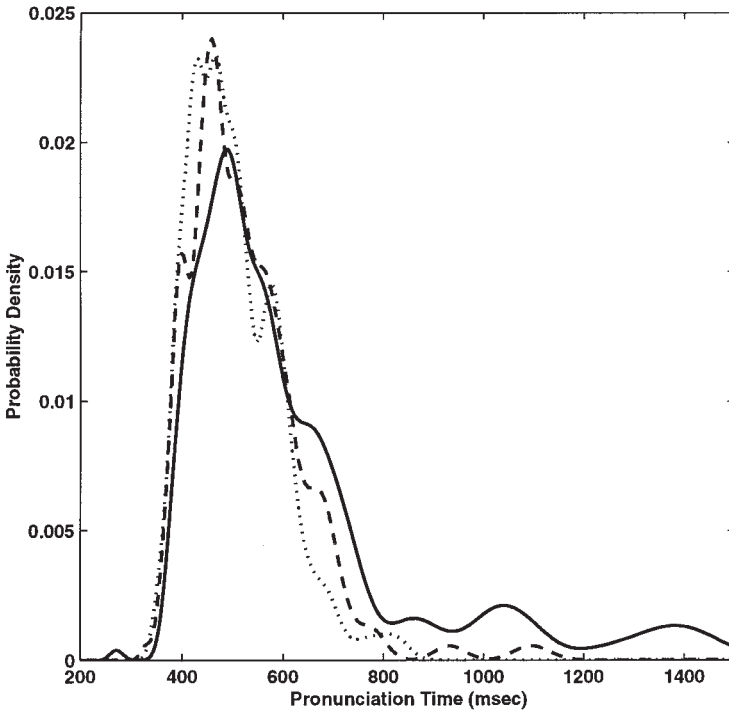
FIGURE 2   The dotted line pertains to the kernel-smoothed density function of pronunciation times measured at the grapheme scale. The dashed line depicts the pronunciation time variability at the spelling-body scale. The solid line represents the pronunciation time variability measured on a whole-word spelling scale. The greater variability at the whole-word spelling scale is apparent in the density function's longer, more stretched tail, as compared to the same measurement taken at the spelling-body scale, which is, in turn, more variable than the same measurement taken at the grapheme scale.

time. Presenting many words to many participants yields a range of sampled control-parameter values. As such, not all word presentations yield identical measurements. Density functions of pronunciation time yield probabilistic estimates of the range of naming-event durations. The visual and statistical analyses identified systematic differences in relative variability in pronunciation time—changes that exceeded the boundaries of change that are likely due to measurement error. Apparently, a wider range of control-parameter values were sampled as the resolution of the measurement scale was increased.

The particular measurement procedures—the measurement scale—used to make contact with the naming act influenced the measurement result. All the measurements, taken using each class of words, each scale, are legitimate measurements. After all, there is no a priori reason to point to any particular scale as a tool that is more or less suitable for taking naming measurements. To consider the pro-

nunciation times to all words, together, would only increase the range of the measurements, which again implies that pronunciation time does not conform to a characteristic measurement scale.

As in Experiment 1, the result of my variability measurement increased as I moved to more detailed measurement scales. Conventional word-recognition theories are silent as to why systematic increases in variability appear as the measurement scale is changed. Conventional perspectives assume that pronunciation times conform to a characteristic scale of measurement. Implicit in the assumption of a characteristic scale of measurement is the requirement that disagreement among results of measurements, taken at different scales, are unsystematic and due only to measurement error. However, a basic method used to characterize a fractal process is to describe precisely how the process interacts with the measurement procedure—to establish a scaling relation.

As predicted, the variability measurement taken at a whole-word scale was amplified over and above the same measurement taken at a spelling-body scale, which was in turn amplified over and above the same measurement taken at a grapheme scale. Moreover, the increasing progressive rank ordering of the variability measurements is consistent with the idea that the process of resolving ambiguity may have no characteristic scale of measurement. The result of the variability measurement appears to depend on where I "step into" the process to take my measurement. After all, other things being equal, a homograph does not entail any *more* grapheme scale vowel ambiguity than a word such as *duck* whose vowel ambiguity is fully resolved at the lower resolution spelling-body scale. Next, I illustrate the same pattern of variability amplification in lexical decision performance using nested ambiguity in relations between pronunciations and patterns of spelling—the mirror image of the ambiguity scale in spelling–pronunciation relations.

EXPERIMENT 3: PHONEME SCALE VERSUS RIME-BODY SCALE AMBIGUITY

The next two experiments use the lexical decision task. Lexical decision measures the time required to decide whether a presented letter string is a legitimate English word. The first lexical decision study is a conceptual replication of Experiment 1. Experiment 1 contrasted variability in pronunciation times for words on a grapheme scale versus a spelling-body scale of ambiguity. Naming performance requires the participant to supply a correct pronunciation; it emphasizes the relation between spelling and pronunciation (Ziegler, Montant, et al., 1997). Lexical decision performance emphasizes the inverse relation, the relation between word pronunciations and spellings (Stone et al., 1997; Ziegler, Montant, et al., 1997; Ziegler, Van Orden, & Jacobs, 1997).

It may not be possible to know all that transpires in making a lexical decision. However, it is intuitive that one must judge whether the presented spelling is that

of a known word. Previously established effects of rime-body ambiguity suggest that spelling verification takes this relation into account. If lexical decision entails something like this, then pronunciations that map to more than one spelling option may amplify variability in performance measures, just as spellings that map to more than one pronunciation do.

This experiment estimates variability in response time using two scales of pronunciation ambiguity. Response-time distributions for words that have invariant rime-body relations (*duck,* hereafter called the phoneme scale) are contrasted with words that have rime-body ambiguous relations (e.g., compare /ûrn/ in *fern* and *turn,* hereafter called the rime-body scale). Again, with respect to the coastline analogy, this experiment contrasts variability estimated on the phoneme scale with variability estimated on the rime-body scale. The move from a phoneme scale to a rime-body scale moves to a more detailed measurement scale, "zooming in" on lexical decision performance.

As with Experiment 1, the method of this experiment was constructed to mimic previous demonstrations of rime-body scale ambiguity. There is some question in the literature as to whether rime-body scale ambiguity reliably influences performance (Peereman, Content, & Bonin, 1998), and an independent replication, which relies on a strong manipulation of rime-body scale ambiguity, may help resolve this debate.

Method

Participants.   Seventy-four introductory psychology students, who were all right-handed native English speakers, participated in the experiment.

Materials.   The key stimuli were 20 yoked pairs of phoneme scale and rime-body scale ambiguous words. *Pronunciation rimes* refer to the particular pronunciations of word bodies (e.g., /ûrn/, in *fern* and *turn*). A *pronunciation neighborhood* is defined as all the monosyllabic words that use that same pronunciation rime. For example, *fern, turn, learn,* and all monosyllabic words that rhyme with *fern* comprise /ûrn/'s pronunciation neighborhood. Ambiguous rime-body relations are pronunciation rimes, such as /ûrn/, that map to more than one spelling body. The *spelling ratio* provides a quantitative estimate of the relative dominance of a rime-body relation that is based on the word frequencies of words that either share or do not share particular rime-body relations in a pronunciation neighborhood. The spelling ratio of a rime-body relation is estimated by dividing the sum of the frequencies of words that share a rime-body relation into the overall sum of the frequencies of all words that have the pronunciation rime. If, in cases where there are exactly two possible spellings, this ratio is less than one half, the rime-body relation is a subordinate relation. In the case where there are several alternative rime-body relations in a pronunciation neighborhood, the subordinate relation is the relation with the smallest spelling ratio.

I only used rime-body scale ambiguous words that have subordinate rime-body relations, and I used no items that had whole-word dominant spelling ratios established by a single high-frequency word. Phoneme scale items that have large spelling neighborhoods with several high-frequency members were preferred in the selection process. For example, *duck* has several high-frequency neighbors (e.g., *truck, luck,* and *stuck*). Hermit words that are themselves the only member of a spelling neighborhood were not used as targets. Five rime-body invariant relations appeared more than once in the experiment.

Only two of the items with invariant rime-body relations (*rape* and *click*) had alternative spelling-to-pronunciation relations with frequencies above zero per million. The exception spelling bodies are rare: *Crepe* has a frequency of one, and *chic* has a frequency of seven. All of the items chosen for the rime-body scale stimulus list were strictly body-rime invariant. This was important because spelling-body scale ambiguity affects lexical decision performance (Stone et al., 1997).

The yoked item pairs were closely matched for word frequency: Phoneme Scale, $M = 3.70$ ($SD = 2.60$); Rime-Body Scale, $M = 3.80$ ($SD = 2.31$); and number of letters: Phoneme Scale, $M = 4.45$ ($SD = 0.51$); Rime-Body Scale, $M = 4.50$ ($SD = 0.51$). The experimental stimuli were embedded in a list of 138 filler words and 178 (half four-letter, half five-letter) pronounceable nonwords, drawn at random from a nonword database supplied by Greg Stone. The yoked item pairs appear in Appendix C.

Procedure.   A trial began with a fixation stimulus $(+++)$ presented for 415 msec at the center of a standard PC screen. The fixation was followed by a blank screen for 200 msec, after which a target letter string appeared. Participants were instructed to respond quickly and accurately by pressing one of two keys; the *yes* key signaled a word letter string, and the *no* key signaled a nonword letter string. The letter string remained visible until the participant responded. An intertrial interval of 629 msec followed the key press. Each participant completed 50 practice trials immediately prior to the 356-trial experimental block, half of which were nonwords. The yokes containing repeated rime-body relations were separated by presenting one yoked pair, embedded among nonwords and fillers, at the beginning of the experiment, and the other pair, also embedded in nonwords and fillers, at the end of the experiment. The presentation order was counterbalanced across participants; post hoc analyses failed to detect a rime-body repetition effect. The remaining targets were distributed evenly among fillers and nonwords and presented in a pseudorandom order. The entire experimental procedure required approximately 20 min to complete.

Results and Discussion

Standard means analysis.   I include a standard participant- and item-means analysis because there is some question in the literature as to whether rime-body

scale ambiguity reliably affects lexical decision performance (Peereman et al., 1998). This analysis represents the typical manner in which the literature presents rime-body scale ambiguity as an "effect" or a difference in means. I wanted to use data from as many participants as possible, although ensuring that the included participants were engaged in the task. The data for 13 of the 74 participants were eliminated because their combined error rates to filler word trials and nonword foil trials exceeded 10%. All analyses were conducted on the response-time data for the remaining 61 participants. (Identical analyses, using stricter and looser exclusionary criteria, yielded the same qualitative pattern of results.) A standard participant- and item-means analysis was conducted on yoked correct response times that fell within the interval between 200 and 2,500 msec. This analysis revealed a reliable effect of rime-body scale ambiguity, which replicates the standard effect, $t_p(60) = 8.70, p < .05$, and $t_i(19) = 4.17, p < .05$. Mean response times to words on a phoneme scale, such as *duck*, were faster than mean response times to words on a rime-body scale, such as *fern*: Phoneme Scale, M = 654.37 (SD = 89.50); Rime-Body Scale, M = 739.37 (SD = 107.24).

Variability analysis.   A pattern of stretching is established if the average difference between the third quartiles of the phoneme- and rime-body scale distributions is reliably larger than the average difference between the same two distribution's first quartiles. I recorded first and third quartiles of each participant's distribution of correct phoneme scale and rime-body scale response-time distributions. I then conducted the stretching analysis, as described in Experiment 1, by first testing for a reliable interaction of the average first- and third-quartile response-time measurements, across the two scales. I conducted two 2×2 repeated measures ANOVAs; the first treated participants as a random variable and the second treated items as a random variable. The interaction term for the participant analysis was statistically reliable, $F_p(1, 60) = 23.77, p < .05$, as was the interaction term for the item analysis, $F_i(1, 19) = 8.57, p < .05$.

Planned contrasts revealed the pattern of the interaction. The average difference of 47 msec between first quartiles of the participant-analysis phoneme and rime-body scale distributions was statistically reliable, $t_p(60) = 7.43, p < .05$, and the average difference of 123 msec between the third quartiles of the same two distributions was larger and also statistically reliable, $t_p(60) = 7.98, p < .05$. Likewise, the average difference of 60 msec between first quartiles of the item-analysis phoneme and rime-body scale distributions was statistically reliable, $t_i(19) = 4.55, p < .05$, and the average difference of 159 msec between the third quartiles for the same two distributions was larger and also statistically reliable, $t_i(19) = 3.77, p < .05$. As predicted, the variability measurement resulting from using a higher resolution rime-body scale is larger than that found at the phoneme scale. A visual contrast of the probability density functions of these distributions, presented in the following figure, corroborates the results of the statistical analysis.

Probability density functions for the two distributions, constructed as described in Experiment 1, are plotted in Figure 3. The dashed line depicts the variability of the distribution of response times to phoneme scale words. The solid line depicts the density function representing the variability found at the rime-body scale. The qualitative pattern revealed in this contrast is virtually identical to the pattern revealed by the parallel contrast at the grapheme and spelling-body scales in naming. The distribution at the rime-body scale has a more stretched tail than the distribution at the phoneme scale. As predicted, the result of my variability measurement is amplified as I move to a higher resolution measurement scale. Next, I describe a nearly identical lexical decision experiment that included contrasts of all three scales of spelling ambiguity.
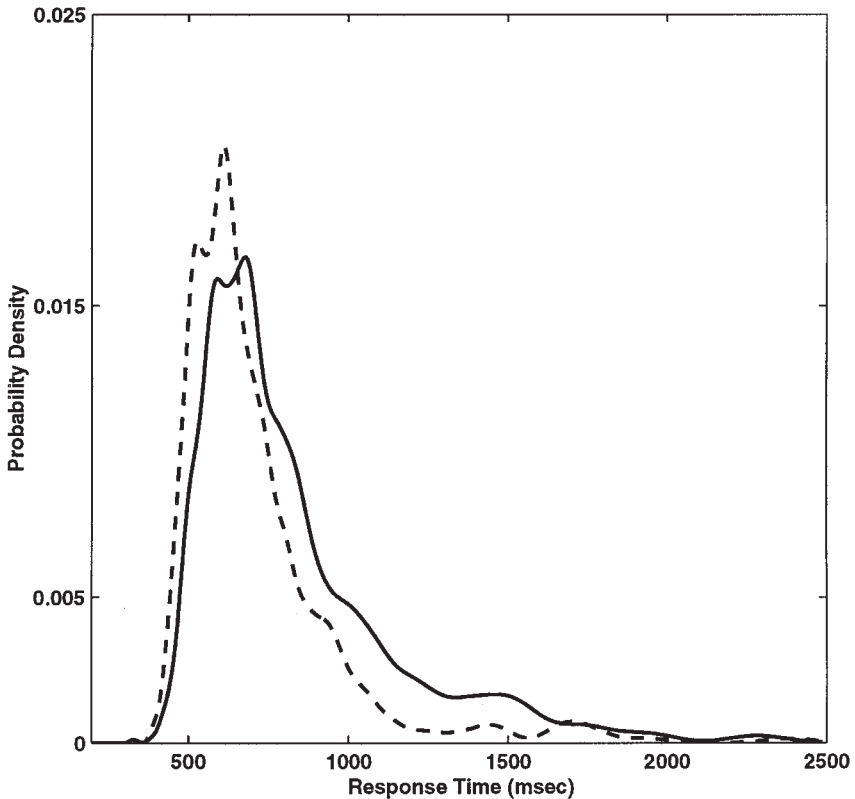

FIGURE 3   The kernel-smoothed probability density functions depict the response-time variability to words at the phoneme scale and rime-body scales. The dashed line depicts the variability measurement at the phoneme scale. The solid line depicts the variability measurement at the rime-body scale. The qualitative stretching pattern revealed by this contrast is virtually identical to the pattern revealed by the contrast at the grapheme and spelling-body scales in naming.

EXPERIMENT 4: PHONEME VERSUS RIME-BODY VERSUS WHOLE-WORD PRONUNCIATION AMBIGUITY

This final lexical decision study is a conceptual replication of Experiment 2, which contrasted pronunciation time variability on a grapheme scale, a spelling-body scale, and a whole-word spelling scale. Homographs have more than one legitimate pronunciation and thus entail all three nested scales of ambiguous spelling–pronunciation relations: grapheme-phoneme, body-rime, and whole-word spelling. Similarly, homophones refer to different words with different meanings but identical pronunciations. Because they are whole-word pronunciation ambiguous, they entail all three nested scales of ambiguous pronunciation–spelling relations.

Experiment 4 measures variability in response-time distributions on a phoneme scale (*duck*), on a rime-body scale (*fern*), and on a whole-word scale (using homophones such as *reed,* hereafter called the whole-word pronunciation scale). I expect that the variability measurement, taken at the whole-word pronunciation scale will be larger than what one would expect, given their status as rime-body scale words. Detecting a variability increase in a contrast between the rime-body and whole-word pronunciation scales requires yoking whole-word pronunciation scale words with rime-body scale words that have similar spelling and pronunciation ratios (see Method). Previous research has established that homophones yield exaggerated lexical decision times compared to nonhomophonic words (Pexman, Lupker, & Jared, 2001).

I conducted Experiment 3 to establish a stronger variability contrast between the phoneme and rime-body scales than I could accomplish in the three-scale contrast of this experiment. Both experiments used the same participants, which meant I could not reuse any phoneme scale words in the direct replication of the contrast at the phoneme and rime-body scales provided by Experiment 4; I gave priority to the contrast at the phoneme and rime-body scales in Experiment 3. Thus, the phoneme scale items used in Experiment 3 came from more common spelling neighborhoods (as measured by adding the frequency counts of all the monosyllabic words that share a phoneme scale word's rime-body relation) than did the phoneme scale words used in Experiment 4. This may imply that weaker historical sources of constraint, and consequent internalized constraints, are available to the process that ends in a lexical decision and thus implies a weaker contrast at the phoneme and rime-body scales in this experiment than in Experiment 3.

Method

Participants.   The participants were the same 74 right-handed, native English-speaking introductory psychology students who participated in Experiment 3.

Materials. The key stimuli were 20 yoked triplicates of phoneme scale (*duck*), rime-body scale (*fern*), and whole-word pronunciation scale (*reed*) words. Only subordinate homophones were used. A *subordinate homophone* is defined as having the lowest frequency count of all the alternative spellings. Thus, *blew*, which has a frequency of 12 per million is subordinate to *blue*, with a frequency of 143 per million.

Word bodies refer to particular spelling patterns (e.g., _int in *mint* and *pint*). A spelling neighborhood is defined as all the monosyllabic words that use that same spelling body. For example, *pint, mint, lint*, and all other words ending in _int comprise _int's spelling neighborhood. Ambiguous body-rime relations are spelling bodies, such as _int, that map to more than one pronunciation rime. The pronunciation ratio provides a quantitative estimate of the relative dominance of a body-rime relation. The pronunciation ratio of a body-rime relation is computed by dividing the sum of the frequencies of words that share a body-rime relation into the overall sum of the frequencies of all the words that have the spelling body.

Similarly, pronunciation rimes refer to the particular pronunciations of word bodies (e.g., /ûrn/, in *fern* and *turn*). A pronunciation neighborhood is defined as all of the monosyllabic words that use that same pronunciation rime. For example, *fern, turn, learn*, and all other monosyllabic words that rhyme with *fern* comprise /ûrn/'s pronunciation neighborhood. The spelling ratio of a rime-body relation is estimated by dividing the sum of the frequencies of the words that share a rime-body relation into the overall sum of the frequencies of all the words that have the pronunciation rime. The rime-body scale and whole-word pronunciation scale yokes were matched with respect to their pronunciation and spelling ratios: Rime-Body Scale pronunciation ratio, $M = .90$ ($SD = .30$); Whole-Word Pronunciation Scale pronunciation ratio, $M = .91$ ($SD = .26$); Rime-Body Scale spelling ratio, $M = .13$ ($SD = .18$); Whole-Word Pronunciation Scale spelling ratio, $M = .16$ ($SD = .23$).

The yoked word triplicates were closely matched for word frequency: Phoneme Scale, $M = 3.95$ ($SD = 3.03$); Rime-Body Scale, $M = 3.80$ ($SD = 2.84$); Whole-Word Pronunciation Scale, $M = 4.05$ ($SD = 3.36$); and length: Phoneme Scale, $M = 4.35$ ($SD = 0.49$); Rime-Body Scale, $M = 4.10$ ($SD = 0.85$); Whole-Word Pronunciation Scale, $M = 4.45$ ($SD = 0.51$). The 60 experimental stimuli were embedded in a list of 118 filler words, and 178 (half four-letter, half five-letter) pronounceable nonwords, drawn at random from a nonword database supplied by Greg Stone. The yoked triplicates appear in Appendix D.

Procedure. The procedures were identical to the those used in Experiment 3.

Results and Discussion

Variability analysis. I recorded the first and third quartiles of each participant's distribution of correct phoneme scale, rime-body scale, and whole-word pronunciation scale response-time distributions. I began the stretching analyses by

first testing for a reliable interaction of the first and third quartile response-time measurements on the phoneme and rime-body scales.

A pattern of increasing variability in a contrast between the phoneme and rime-body scales is established if the difference between the distribution's third quartiles is larger than the difference between their first quartiles. A test of the interaction terms for both the participant and item analyses, using 2×2 repeated measures ANOVAs, failed to reach statistical significance.

Nevertheless, I conducted the planned contrasts to determine if the qualitative pattern of the average quartile differences was consistent with the hypothesis of increasing variability, and it was. The average difference of 17 msec between first quartiles of the participant-analysis phoneme and rime-body scale distributions was statistically reliable, $t_p(60) = 2.59$, $p < .05$, and the average difference of 32 msec between the third quartiles of the same distributions was larger and statistically reliable, $t_p(60) = 1.79$, $p < .05$. The average difference of 9 msec between first quartiles of the item-analysis phoneme and rime-body scale distributions was not statistically reliable, however. The average difference of 34 msec between the third quartiles of the same two distributions was indeed larger but also failed to reach statistical significance. As noted, the phoneme and rime-body contrast was weaker in this experiment than in Experiment 3. The apparent stretching pattern revealed by the planned contrasts mirrors that of the statistically reliable stretching pattern established in Experiment 3, which included a stronger contrast between the phoneme and the rime-body scales.

The critical contrast of Experiment 4 concerns the variability measurements taken at the rime-body and the whole-word pronunciation scales. A test of the interaction term for the participant analysis, using a 2×2 repeated measures ANOVA, at the rime-body and whole-word pronunciation scales, was statistically reliable, $F_p(1, 60) = 26.18$, $p < .05$. Likewise, the interaction term of a 2×2 repeated measures ANOVA, using items as a random variable, was statistically reliable, $F_i(1, 19) = 6.98$, $p < .05$.

The 41-msec average difference between the first quartiles of the participant-analysis rime-body and whole-word pronunciation scale distributions was statistically reliable, $t_p(60) = 4.12$, $p < .05$, and the third quartiles of the same two distributions yielded a larger and statistically reliable average difference of 139 msec, $t_p(60) = 6.94$, $p < .05$. Similarly, the average difference of 58 msec between first quartiles of the item analysis rime-body scale and whole-word pronunciation scale distributions was reliable, $t_i(19) = 2.50$, $p < .05$, and the larger average difference of 176 msec between the third quartiles of the same two distributions was also statistically reliable, $t_i(19) = 3.23$, $p < .05$. Once again, the results of the variability measurement increased, as the move to a higher resolution measurement scale was made. I corroborated the results of the statistical analyses with a visual contrast of the probability density functions of all three distributions.

The overall prediction was a progressive increase in variability measurements that tracks the successively nested, higher resolution scales. Probability density functions for the three distributions, generated as previously described, are plotted
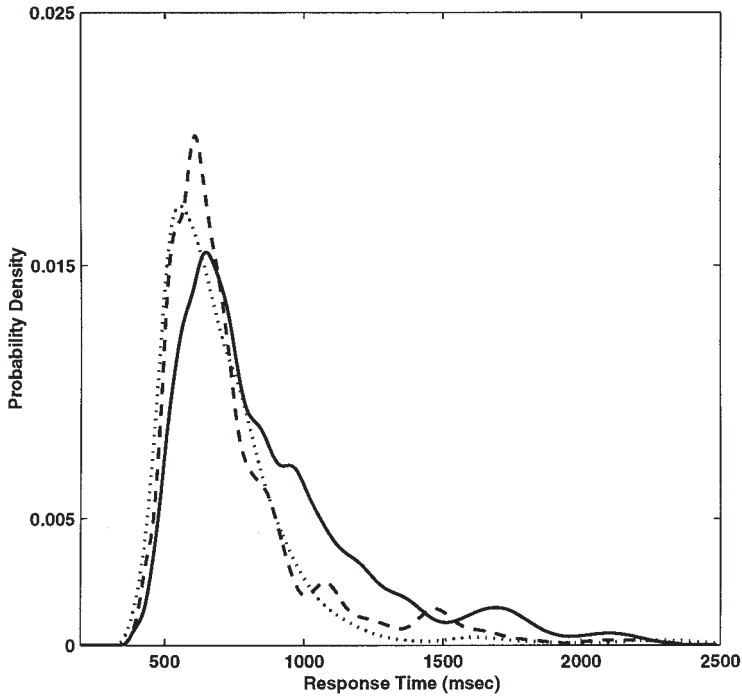
FIGURE 4   The probability density functions to the three scales of ambiguous pronuncia-tion–spelling relations are depicted. The dotted line represents the kernel-smoothed density function of response times to words at the phoneme scale. The dashed line depicts the distribu-tion of response times to words at the rime scale. The solid line depicts the distribution of re-sponse times, now on a whole-word pronunciation scale. The qualitative pattern revealed by this contrast is virtually identical to patterns revealed by the contrast at the grapheme, spell-ing-body, and whole-word scales in naming.

in Figure 4. The dotted line depicts the variability measurement at the phoneme scale. The dashed line depicts the variability found at the rime-body scale. The solid line depicts the result of the variability measurement, now at the whole-word pronunciation scale. The modes of the distributions tend to be similar, but the dis-tributions diverge in their slow tails to include more and more slow responses. As predicted, the result of the variability measurement at each scale is different—it in-creases as scale resolution increases.

INVERSE POWER-LAW BEHAVIOR

Pronunciation and response-time distributions that resemble inverse power-law distributions render plausible the hypothesis that naming and lexical decision are indeed fractal processes. Next, I explore this possibility. If naming and lexical deci-

sion are fractal processes, it may be possible to observe empirical patterns that are more strictly associated with fractal processes.

How might distributions of response time appear if one backed away from all the typical assumptions that go into a conventional means analysis? For instance, let us ignore the a priori biases that distinguish word-response times from nonword-response times, correct response times from incorrect response times, skilled from unskilled performance, and typical responses from atypical responses (i.e., outliers). That is, let us take a very conservative look at how response times are distributed.

Panel A of Figure 5 represents every response I collected that fell between 200 and 5,500 msec for both lexical decision experiments described earlier (i.e., Experiments 3 and 4). This distribution is comprised of more than 26,000 response times. It includes both *correct* and *incorrect* times to all targets, all fillers, and all nonword foils. It includes the responses from all 74 participants, even those whose data were originally excluded because they failed to meet my accuracy criterion.

The distribution's mode is 600 msec. Past that point, the distribution's slow tail diminishes dramatically, but nevertheless includes responses even in the neighborhood of 5 sec. This distribution appears as a unitary object, even though it represents responses to many different words and nonwords taken from many different participants.

Panel B of Figure 5 displays the same probability density function, now plotted on double logarithmic coordinates. In this domain the *x* axis corresponds to the natural logarithm of response time and the *y* axis corresponds to the natural logarithm of the probability density. Apparently, the slow tail of the probability density
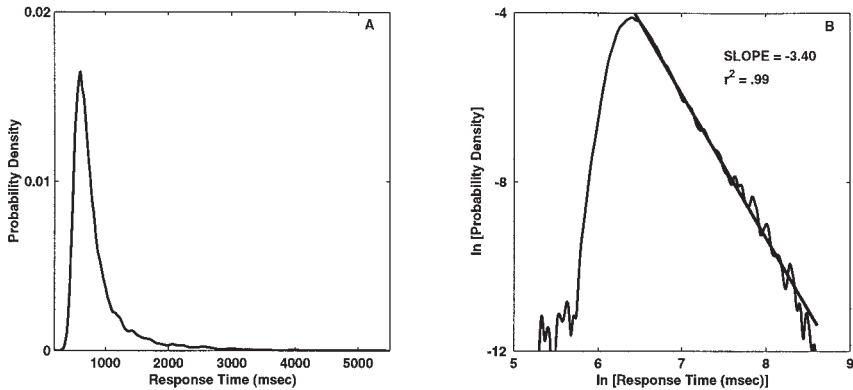
FIGURE 5   Panel A depicts the kernel-estimated density for all responses from Experiments 3 and 4 that fell between 200 and 5,500 msec. Included are both correct and incorrect responses to both words and nonwords for all 74 participants. Panel B depicts the same density, now plotted on double logarithmic coordinates. The heavy solid line represents a least squares regression line, fit in the logarithmic domain, over the interval extending from the density's mode to the end of its slow tail. Apparently, the distribution's slow tail is well described by an inverse power-law scaling relation.

diminishes as a linear function of response time in logarithmic coordinates, with a slope of –3.40 ($r^2 = .99$); the wavy oscillations are an artifact of the kernel procedure estimating very sparse intervals of the distribution.[1]

An inverse power-law scaling relation is precisely a linear relation on double logarithmic coordinates along the slow tail of a distribution (Jensen, 1998; West & Deering, 1995). This represents a positive, conservative test for a scaling relation in lexical decision performance. Apparently, the probability density function of response time is well described in terms of an inverse power-law scaling relation—the line extending from the distributions mode out to the end of its slow tail. This is the strongest evidence I have presented so far that response times may have no characteristic scale of measurement.

From this perspective, even the very slow 5-sec response times obey the same power-law scaling relation as the faster response times. Thus, the relative number of very slow response times is strictly related to the relative number of very fast response times, and vice versa. A conventional perspective presumes that different classes of words may require different mechanisms (e.g., Coltheart, 1978; Coltheart, Curtis, Atkins, & Haller, 1993), which are revealed by distinct time courses. Alternatively, the fractal perspective allows that all response times gauge the self-same process. Here, all the response times closely follow a power law, and no natural break point exists that would distinguish additional processes.

I have speculated that the increases in variability described in all four experiments result from the higher resolution scales increasing their access to power-law behavior. They reveal a more exhaustive sample of the range of possible performances. If so, then classic cognitive assumptions may simply restrict what is admitted into the measurement process—the resolution or entry point from which a scientist may observe a cognitive performance—just as one might restrict which sub-bays and subpeninsulas are admitted into measurements of the coastline of Great Britain.

The introduction illustrated how printed text, a cultural artifact, resembles a fractal pattern (Van Orden et al., 2001), and performance yields patterns of variability that are consistent with fractal processes. The experiments demonstrated how the fractal structure of the cultural artifact constrains performance, and ultimately (over much slower time scales of cultural change) the fractal patterns of performance supply constraints to the "trajectory" of the cultural artifact (e.g., the fluctuating patterns of typical spellings and pronunciations; cf. *wast* and *was*; see also Van Orden, Moreno, & Holden, 2003).

Perhaps it is here that fractal perspectives on performance make contact with ecological perspectives on performance. Iterative models of reading performance

---

[1]An inverse power-law density function that has a slope of –2 or greater has infinite variance, and a slope of –1 or greater implies an infinite mean (Jensen, 1998). There were too few pronunciation times to present them in an identical plot. Elsewhere, my colleagues and I established the same pattern in pronunciation time distributions (Holden, Van Orden, & Turvey, 2002).

formally illustrate how a fractal perspective implies that linguistic performances entail irreducible relations between the linguistic environment and embodied linguistic constraints—constraints that emerge as a consequence of the idiosyncratic details of a person's history of interaction in his or her linguistic environment. Likewise, ecological perspectives seek to catalog the irreducible relations that constrain other kinds of perceptual performances. For example, haptic perceptions do not simply gauge an isolated property of an object, such as its mass. Instead, they entail irreducible relations between an object's objective physical properties and the properties of the force-producing neuromuscular system that is wielding the object (Turvey, Shockley, & Carello, 1999)—irreducible relations among environmental and embodied constraints.

SUMMARY

In each of the four experiments I presented, variability measurements for pronunciation and lexical decision times changed as a function of the scale of resolution used to take the measurement. These experiments corroborate the hypothesis that variability in response time is amplified as the scale of ambiguity accesses nested ambiguity in the relations between word spellings and pronunciations. Furthermore, the response times from the two lexical decision experiments were shown to obey an inverse power-law scaling relation, which suggests response time does not conform to a characteristic measurement scale.

So what kinds of systems have an inherent capacity to produce the kinds of systematic changes in variability that were observed here? As noted earlier, the same process repeating itself can be viewed as iteration in time, as an iterative dynamical system. Some processes that give rise to fractal patterns result from the repeated application of a simple mathematical rule. Recurrent network models of word naming (e.g., Farrar & Van Orden, 2001; Kawamoto & Zemblidge, 1992) and word recognition (e.g., Gibbs & Van Orden, 1998) are also examples of iterative processes that are predicated on multiplicative interaction.

The approximate lognormal shape of the grapheme scale pronunciation time distribution suggests multiplicative interaction of system (control) parameters. Naming is a laboratory task that more closely parallels normal reading activities than lexical decision, for example (Bosman & Van Orden, 1997). As such, it may enlist well-tuned, well-learned relations between spellings and their pronunciations, resulting in relatively stable performances. If so, variability distributed in the lognormal shape may be a benchmark of skilled performance (cf. Holden, Van Orden, & Turvey, 2002). Nearly all the other response-time distributions diverged from the lognormal shape, with exaggerated slow tails. Thus, most response-time distributions better resemble *inverse power-law distributions*, distributions in which variability outpaces the benchmark established by the more constrained multiplicative model (Schroeder, 1991; West & Deering, 1995).

In my introduction I described how recurrent networks represent a model that describes how variability may accumulate in a multiplicative iterative process. In a fractal process, however, the amplification of variability outpaces the limits of multiplicative interaction, and parametric measurements of variability will entail even more dramatic, proportional amplification, which is observed as a power-law scaling relation.

Variability measurements at higher resolutions in the nested fractal relations between English spellings and pronunciations were larger than the same measures of variability taken with lower resolution scales in both the word naming and lexical decision tasks. Apparently, a wider range of iterations in the process that resolves ambiguity in spelling–pronunciation relations was revealed as higher resolution ambiguity scales were introduced to the measurement procedure. In terms of my analogy to the measurement of a coastline, it was as if more jagged details of the coastline were revealed as more detailed measurements were made.

ACKNOWLEDGMENTS

REFERENCES

Bassingthwaighte, J. B., Liebovitch, L. S., & West, B. J. (1994). *Fractal physiology.* New York: Oxford University Press.

Bosman, A. M. T., & Van Orden, G. C. (1997). Why spelling is more difficult than reading. In C. A. Perfetti, L. Rieben, & M. Fayol (Eds.), *Learning to spell* (pp. 173–194). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

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

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

Farrar, W. T., IV, & Van Orden, G. C. (2001). Errors as metastable response options. *Nonlinear Dynamics, Psychology, and Life Sciences, 5,* 223–265.

Gibbs, P., & Van Orden, G. C. (1998). Pathway selection's utility for control of word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1162–1187.

Glushko, R. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance, 2,* 361–379.

Holden, J. G., Van Orden, G. C., & Turvey, M. T. (2002). *Inverse power-law behavior and response time distributions.* Manuscript submitted for publication.

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

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

Jensen, H. J. (1998). *Self-organized criticality.* Cambridge, England: Cambridge University Press.

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

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

Mandelbrot, B. B. (1983). *The fractal geometry of nature.* San Francisco: Freeman.

Peereman, R., Content, A., & Bonin P. (1998). Is perception a two-way street? The case of feedback consistency in visual word recognition. *Journal of Memory and Language, 39,* 151–174.

Pexman, P. M., Lupker, S. J., & Jared, D. (2001). Homophone effects in lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 139–156.

Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin, 3,* 446–461.

Ratcliff, R., & Murdock, B. B. (1976). Retrieval processes in recognition memory. *Psychological Review, 83,* 190–214.

Riley, M. A., & Turvey, M. T. (in press). Variability and determinism in motor behavior. *Journal of Motor Behavior.*

Schroeder, M. R. (1991). *Fractals, chaos, power laws: Minutes from an infinite universe.* New York: Freeman.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis.* London: Chapman & Hall.

Spieler, D. H., & Balota, D. A. (1997). Bringing computational models of word naming down to the item level. *Psychological Science, 8,* 411–416.

Stone, G. O., Vanhoy, M. D., & Van Orden, G. C. (1997). Perception is a two-way street: Feedforward and feedback phonology in visual word recognition. *Journal of Memory and Language, 36,* 337–359.

Turvey, M. T., Shockley, K., & Carello, C. (1999). Affordance, proper function, and the physical basis of perceived heaviness. *Cognition, 73,* B17–B26.

Ulrich, R., & Miller, J. (1993). Information processing models generating lognormally distributed reaction times. *Journal of Mathematical Psychology, 37,* 513–525.

Van Orden, G. C., Moreno, M. A., & Holden, J. G. (2003). A proper metaphysics for cognitive performance. *Nonlinear Dynamics, Psychology, and Life Sciences, 7,* 47–58.

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

Van Orden, G. C., Pennington, B. F., & Stone, G. O. (2001). What do double dissociations prove? *Cognitive Science, 25,* 111–172.

Van Zandt, T. (2000). How to fit a response time distribution. *Psychonomic Bulletin & Review, 7,* 424–465.

West, B. J., & Deering, B. (1995). *The lure of modern science: Fractal thinking.* River Edge, NJ: World Scientific.

Ziegler, J. C., Montant, M., & Jacobs, A. M. (1997). The feedback consistency effect in lexical decision and naming. *Journal of Memory and Language, 34,* 567–593.

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

APPENDIX A
Yoked Grapheme Scale and Spelling-Body Scale Items

Grapheme Scale Item	Number of Letters	Word Frequency	Spelling-Body Scale Item	Number of Letters	Word Frequency
BOLT	4	10	BUSH	4	14
SLANG	5	2	SOOT	4	1
BARGE	5	7	BROOD	5	9
WEDGE	5	4	WHARF	5	4
DICE	4	14	DOME	4	17
WICK	4	4	WARP	4	4
SKID	4	2	SWAP	4	2
SURF	4	1	SWATH	5	1
WISP	4	2	WASP	4	2
PRANK	5	1	PLAID	5	1
ROBE	4	6	ROUGE	5	7
SPOIL	5	3	SKULL	5	3
QUAKE	5	2	QUART	5	3
SCRUB	5	9	SQUAT	5	7
DUCK	4	9	DOLL	4	11
BUDGE	5	3	BROTH	5	3
HUMP	4	2	HOOF	4	2
PILL	4	15	PINT	4	13
WELD	4	4	WORM	4	4
SNAG	4	3	SWAN	4	3
WEED	4	1	WAND	4	1
M	4.38	4.95		4.43	5.33
SD	0.5	4.19		0.51	4.78

APPENDIX B
Yoked Grapheme, Spelling Body, and Whole-Word Spelling Scale Items

Grapheme Scale Item	Word Frequency	Dominant Spelling Body Scale Item	Dominant Pronunciation Ratio	Word Frequency	Subordinate Spelling-Body Scale Item	Subordinate Pronunciation Ratio	Word Frequency	Whole-Word Spelling Scale Item	Subordinate Pronunciation Ratio	Word Frequency
WAKE	23	WASTE	.97	35	WARD	.08	25	WOUND	.02	28
PUMP	11	PEAK	.66	10	PORK	.04	10	BASS	.03	16
WAGE	56	WOOD	.76	56	WAVE	.11	46	WIND	.05	63
TUNE	10	TOSS	.72	9	TOMB	.22	11	TEAR	.06	11
KEPT	186	GIVE	.57	391	COST	.24	229	CLOSE	.15	234
DIME	4	TOAD	.75	4	DRONE	.07	3	DOVE	.19	4
LACK	110	LIST	.68	133	LAID	.17	77	LEAD	.21	129
M	57.14		.73	91.14		.13	57.29		.1 (.9)	69.29
SD	68		.12	139.63		.08	79.97		.08	84.60

APPENDIX C
Yoked Phoneme Scale and Rime-Body Scale Items

Phoneme Scale Item	Number of Letters	Word Frequency	Rime-Body Scale Item	Number of Letters	Word Frequency
RAPE	4	5	RUDE	4	6
HUMP	4	2	HOOP	4	3
FROG	4	1	FERN	4	1
SPIT	4	11	SPED	4	9
PUKE	4	1	POMP	4	3
GULP	4	2	GRAIL	5	2
BARGE	5	7	SEIZE	5	6
STING	5	5	SMIRK	5	3
LOFT	4	2	YAWN	4	2
ROBE	4	6	STIR	4	7
KELP	4	2	CRUX	4	2
HUNCH	5	7	FRAUD	5	8
NUDGE	5	2	NICHE	5	3
HULK	4	2	HURL	4	3
LUST	4	5	MULE	4	4
PUNCH	5	5	TRUCE	5	5
SLANG	5	2	STUNG	5	2
CLICK	5	2	GAUZE	5	1
BUDGE	5	3	BOOZE	5	4
FLING	5	2	PLUME	5	2
M	4.45	3.70		4.50	3.80
SD	0.51	2.60		0.51	2.31

APPENDIX D
Yoked Phoneme, Rime-Body, and Whole-Word Pronunciation Scale Items

Phoneme Scale Item	Word Frequency	Rime-Body Scale Item	Word Frequency	Pronunciation Ratio	Spelling Ratio	Whole-Word Pronunciation Scale Item	Word Frequency	Pronunciation Ratio	Spelling Ratio
COIN	10	HUNT	10	1	.11	BLEW	12	0.99	.07
HOOK	5	BEAN	5	1	.37	REED	5	1	.78
BUST	7	BLAZE	7	1	.06	BEECH	6	1	.06
DRIP	1	BRAWL	1	1	.005	BOAR	1	1	.001
CUBE	1	HONE	2	0.07	.18	HARE	1	0.11	.08
WEDGE	4	BLISS	4	1	.05	BERTH	4	1	.01
SKID	2	SNIFF	2	1	.02	FEINT	2	1	.02
MASK	9	SLUM	8	1	.02	HYMN	9	1	.003
MARSH	5	QUEER	6	1	.04	WEIGH	4	1	0
STUB	3	QUIZ	2	1	0	LIEN	2	1	.003
PROBE	6	RINSE	6	1	.008	REIGN	7	1	.008
DUCK	9	GOAT	6	1	.32	SEAM	9	1	.47
ROACH	2	PLAID	1	0	0	KNEAD	1	0.17	.21
TORCH	2	CLOAK	3	1	.08	CHUTE	2	1	.08
ROOST	1	GLIDE	2	1	.62	GRATE	3	1	.57
SMUG	7	STUD	7	1	.28	RITE	8	1	.26
BLIMP	1	PERCH	1	1	.002	WAIVE	1	1	.002
BRAG	2	PERK	1	1	.05	YOLK	1	1	.11
SURF	1	PUFF	1	1	.45	REEL	2	1	.43
SOCK	2	MOAN	1	1	.03	LUTE	1	1	.08
M	3.95		3.8	0.904	0.135		4.05	0.914	0.162
SD	3.03		2.84	0.297	0.178		3.36	0.265	0.225