

DISTINGUISHING COMMON AND TASK-SPECIFIC PROCESSES IN WORD
IDENTIFICATION: A MATTER OF SOME MOMENT?

Sally Andrews
School of Psychology
University of New South Wales

&

Andrew Heathcote
Department of Psychology
University of Newcastle

Running head: Common and task-specific processes in word identification

Correspondence:

Sally Andrews
School of Psychology
University of New South Wales
Sydney, 2052
Australia
Email: S.Andrews@unsw.edu.au
Fax: 61-2-93853641

Abstract

In Experiment 1, 500 words were presented to 75 participants in five different word identification tasks (lexical decision, semantic categorization and three speeded naming tasks) to investigate differences between the reaction time (RT) distributions, and evaluate the basis of effects of word frequency and animacy, in different tasks. Experiment 2 collected delayed naming latencies for the same words to establish that the observed naming effects were not due to production-related characteristics of the items. Low frequency words yielded more skewed distributions than high frequency words in all tasks except delayed naming, but the differential skew was most marked for tasks that required lexical discrimination. The semantic categorization task yielded highly skewed distributions for all words, but the word frequency effect was due to shifts in the location of the RT distributions rather than changes in skew. The results are used to evaluate the relative contributions of a common lexical access process and task-specific processes to performance in lexical discrimination and naming tasks.

**Distinguishing common and task-specific processes in word identification:
A matter of some moment?**

Visual word recognition is one of the most intensively researched domains in cognitive psychology. Although models and theoretical interpretation have increased in sophistication, the methodologies used to investigate the domain have remained fairly constant (but see Balota & Abrams, 1995) and relatively crude. Research has relied almost exclusively on comparisons of average speed and accuracy of performance for different stimulus classes and/or tasks. Such comparisons may, however, lead to imprecise, or even misleading conclusions about the processes underlying performance (e.g., Miller, 1991). Rather than reflecting a simple change in the location of the RT distribution, changes in mean RT might arise from more complex changes in the distribution's shape that signal the involvement of different underlying processes (Ratcliff & Murdock, 1976; Heathcote, Popiel & Mewhort, 1991). The present research estimated the RT distributions for the same high frequency (HF) and low frequency (LF) words in five different word identification tasks that required binary classification or naming responses. Changes in distribution shape across different tasks provide the basis for a more refined evaluation of the processes involved in visual word recognition than is possible from measures of average RT alone. The research focuses particularly on the effects of word frequency on the RT distributions obtained in different word identification tasks.

Frequency effects on lexical access

Almost all tasks requiring word identification yield faster and more accurate performance for HF than LF words. This "word frequency effect" (WFE) is so ubiquitous that it has determined the lexical retrieval assumptions of all models of lexical access. Serial models attribute it to the ordering of the comparison process (e.g., Becker & Killion, 1977; Forster, 1976) while parallel activation frameworks assume that the representations of HF and LF words have different thresholds (Morton, 1970) or resting activation levels (McClelland and Rumelhart, 1981). Similarly, parallel distributed processing (PDP) models (Seidenberg & McClelland, 1989; Plaut, McClelland,

Seidenberg & Patterson, 1996) assume that connection weights are stronger for more frequently experienced patterns.

This common assumption of a frequency-sensitive lexical retrieval process defines a “conventional” view of the lexical “processing architecture” (Monsell, Doyle & Haggard, 1989, p.45), which assumes that a common set of processes underlie performance on a variety of word identification tasks, and that these common processes usually precede task-specific processes. This view has been caricatured as the “magic moment” approach to lexical access (Balota, 1990) because models often imply that there is a discrete point in time when a lexical representation is “accessed”.

Consistent with this shared assumption, word frequency is the most robust predictor of both lexical decision (e.g., Whaley, 1981) and naming (e.g., Treiman, Mullenix, Bijeljac-Babic & Richmond-Welty, 1995) performance. However, the WFE has consistently been found to be larger in the lexical decision task (LDT) than in naming tasks (e.g., Andrews, 1982; Forster & Chambers, 1973). This implies either that the two tasks do not rely on the same access process, or that frequency effects are due, at least in part, to task-specific processes. There are two alternative interpretations of the differential frequency effects that yield opposite conclusions as to which task provides the more valid index of lexical access.

Naming tasks. Balota and Chumbley (1985) argued that the WFE observed in naming tasks has little to do with lexical access on the basis of data from “delayed naming tasks”, in which production of a target word is withheld until a cue appears. With long enough cue delays, lexical access should be completed before pronunciation begins and any residual WFE must reflect production processes (Forster & Chambers, 1973). Because they found equivalent WFEs in immediate and delayed naming tasks, Balota and Chumbley (1985) concluded that “a large component of the frequency effect in the pronunciation task involves production rather than simple lexical access” (p. 95).

This conclusion severely challenges the utility of the word naming task as an index of lexical access so Balota and Chumbley’s (1985) results were subject to considerable scrutiny. Some investigations have replicated their findings of frequency effects on delayed naming performance (e.g., Balota & Shields, 1988; Connine, Mullenix, Shernoff

& Yelen, 1990), while others have not (e.g, McRae, Jared & Seidenberg, 1990; Monsell et al., 1989; Savage, Bradley & Forster, 1990). Although the reason for these contradictory results remains unclear, the accumulated evidence clearly demonstrates that a WFE can be observed in delayed naming tasks even under conditions that provide sufficient time and “incentive to prepare fully for articulation” (Monsell et al., 1989, p. 50) and for stimuli that are very carefully matched on phonological characteristics (Balota & Shields, 1988; Goldinger, Azuma, Abramson & Jaim, 1997; but see McRae et al., 1990). However, the delayed naming WFE is considerably smaller than that observed in standard speeded naming conditions implying that speeded naming performance does show a reliable WFE that is not due to production. The question still remains as to why this effect is so much smaller than that obtained in the LDT given that both tasks are assumed to engage the same lexical access mechanism.

Dual route models of word naming claim that naming tasks underestimate the WFE due to lexical access because of the influence of nonlexical naming procedures (Coltheart, 1980). In Coltheart et al.'s (1993) computationally implemented Dual Route Cascade (DRC) model, a phonological output buffer receives parallel input from the lexical representations activated by the stimulus and a frequency-insensitive rule algorithm that assigns the phoneme most frequently associated with each grapheme. The rule-based procedure is wrong for exception words such as pint or sword, but generates the correct pronunciation of regular words. The lexical retrieval procedure generates correct pronunciations of all words but is slower for LF than HF words. When the rule algorithm assigns phonemes before lexical access is complete, word frequency will not influence performance. Thus, the reduced WFE in the naming task arises because LF regular words can sometimes be pronounced without completion of the lexical access process that is necessary for a LDT response. Consistent with this interpretation, under conditions that preclude reliance on nonlexical pronunciation assembly, naming tasks yield WFEs equivalent to the LDT (Paap, McDonald, Schvaneveldt & Noel;1987; Monsell et al., 1989; but see Balota & Chumbley, 1990).

Non-lexical influences on pronunciation can also be accommodated by “single-route” frameworks, such as the interactive activation (IA) and PDP models, which do not

assume separate lexical and nonlexical representational systems. Such models assume that the associations between the orthographic and phonological units corresponding to different representational levels (e.g., letters, graphemes, syllables) are determined by the co-occurrence frequency for that unit size (Taft, 1992). Orthographic-phonological associations between lexical units will, therefore, depend on word frequency while those between sub-lexical units will depend on the correspondence frequency of, say, graphemes and phonemes. It is therefore possible to attribute differential frequency effects in naming and LDT tasks to the greater impact of sublexical relationships on pronunciation responses as long as “lexical and sublexical levels have somewhat different time courses” (Monsell et al., 1989, p. 66).

LDT tasks. It might be the LDT rather than the naming task that provides an inaccurate estimate of the role of frequency in lexical access due to contamination from task-specific decision processes. Balota and Chumbley (1984) compared the size of the WFE obtained in the LDT with that in a “semantic categorization” task in which participants judged whether or not a target word was a member of a previously presented semantic category. They found virtually no effect of word frequency on either “yes” or “no” responses in the semantic categorization task, even though the same items yielded substantial frequency effects in the LDT. The absence of frequency effects on “yes” responses task might be due to priming of the target item by the category label, but Balota and Chumbley argued that the lack of a frequency effect on “no” responses “is considerably more difficult to dismiss” (1984, p. 341). Balota and Chumbley concluded that the frequency effects observed in the LDT are due to task-specific processes and may be irrelevant to the processes underlying normal word identification.

Using Atkinson and Juola’s (1974) general two-stage framework for memory retrieval, Balota and Chumbly (1984) elaborated a model in which LDT responses are based on the “familiarity/meaningfulness (FM)” of the target stimulus. Items with extreme FM values (e.g., high frequency words or unwordlike nonwords) yield fast classifications while items with more intermediate values require a second stage of analysis which is assumed to be more attention-demanding and analytic - and therefore slower - than the processes underlying fast first-stage responses: “For example, the

subject may actually need to check the spelling of the letter string against the spelling of the word contained in the subject's lexicon" (Balota & Chumbley, 1984, p. 35). The WFE occurs because LF words have lower FM values than HF words and are therefore more likely to require the slow second-stage analysis.

Balota and Chumbley's claim that frequency does not affect semantic categorization performance remains contentious. Although semantic judgment tasks have been reported to yield smaller WFEs than the LDT (e.g., Forster, 1985; Monsell, 1985), both Monsell et al. (1989) and Forster and Shen (1996) did find WFEs of similar magnitude to those typically found in the LDT in a task requiring binary semantic judgments. Thus, although the precise conditions required to yield a WFE in the semantic categorization task remain unclear, the effect is sometimes as large as the LDT.

Despite these empirical ambiguities, Balota and Chumbley's (1984) emphasis on the role of familiarity and decision processes in the LDT has been influential. Notably, a familiarity mechanism that allows fast non-lexically based decisions in the LDT has recently been incorporated into computational implementations of the two major classes of model of visual word identification: Grainger and Jacobs' (1996) extension of the IA framework - the Multiple Read Out Model (MROM); and the most recent version of the DRC model (Coltheart, Rastle, Perry, Langdon & Ziegler, in press). Both models compute a measure of the summed lexical activity elicited by a stimulus early in processing. If this measure exceeds a threshold, a fast "word" classification can be made even if no word node has yet exceeded threshold. With respect to the WFE, this account is similar in spirit to the first stage of Balota and Chumbley's decision mechanism because it assumes that LDT responses can sometimes be made on the basis of overall stimulus familiarity without unique item identification. HF words have higher resting activation levels and will on average, therefore, generate higher levels of summed lexical activity early in processing than LF words. Consequently, HF words may be classified as words on the basis of summed lexical activity before the individual word node exceeds threshold. LF words with many lexical neighbors may also lead to sufficiently high levels of overall activity to exceed the "word" response threshold (Andrews, 1997), but LDT responses to most LF words require unique identification¹.

There is an important conceptual difference between Balota and Chumbley's (1984) proposal and the familiarity mechanism implemented in MROM and DRC. Balota and Chumbley's (1984) model assumes that familiarity effects on decision processes in the LDT are independent of lexical retrieval. Monsell et al. (1989) pointed out that accepting that LDT frequency effects are magnified by decision processes does not entail rejecting the view that they arise from the frequency-sensitivity of lexical retrieval processes. A familiarity estimate could be derived by "continuous global monitoring of levels or rates of change of activation during [lexical] identification" (Monsell et al., 1989, p. 67). By this view, rather than providing evidence against the contribution of word frequency to lexical identification, "the frequency sensitivity of the decision process is...one reflection ...of the frequency sensitivity of identification" (Monsell et al., 1989, p. 67). The familiarity mechanisms implemented in MROM and DRC is of the form suggested by Monsell et al. (1989) in that it relies on lexical activity rather than an independent familiarity mechanism. Nevertheless, as in Balota and Chumbley's (1984) model, use of the familiarity metric is assumed to be specific to the LDT.

Beyond measures of central tendency

The evidence reviewed so far has relied on mean RT to index the effects of psycholinguistic variables on task performance. Interpretations based on mean RT data often imply that differences between conditions reflect shifts along the response time scale rather than changes in the shape of the RT distribution. But such assumptions may be unjustified (see Balota & Spieler, 1999, Figure 1). An increase in the proportion of slow RTs (reflected in a more positively skewed distribution) will lead to an increase in mean RT without an overall shift in the mode of the distribution. Conversely, mean RT may stay constant despite major changes in the distribution of RTs. For example, the typical finding of weak or null facilitation effects for congruent stimuli in the Stroop color naming task occurs because a decrease in modal RT compared to a neutral condition is accompanied by an increase in the skew of the RT distribution (Heathcote et al., 1991; Spieler Balota & Faust, 1996).

Part of the reason that researchers have continued to rely on measures of central tendency is that traditional methods for estimating distribution shape are neither efficient

nor robust (Ratcliff, 1979)². Estimates of higher cumulants obtained from the usual sample formulae are extremely sensitive to outliers and provide very variable estimates, so that many thousands of observations are required to achieve sufficient precision. Ratcliff and Murdock (1976) suggested an alternative approach, which assumes a particular underlying probability distribution and then uses a maximum likelihood criterion to estimate values of the distribution's parameters that best fit the data. Unlike sample cumulant estimates, maximum likelihood estimates are sufficiently precise with moderately sized samples (Cox & Hinckley, 1974). However, as well as requiring computationally intensive optimization procedures, the major drawback of this method is that it requires selection of a particular theoretical probability distribution. A number of different functions for RT distributions have been proposed (e.g., Luce, 1986), but most recent investigations have used the ex-Gaussian distribution, which Ratcliff (1979) and others (e.g., Heathcote et al., 1991; Hockley, 1984; Mewhort et al., 1992) have demonstrated to yield a good fit with empirical data from a range of choice RT paradigms.

The ex-Gaussian distribution (Hohle, 1965; Burbeck & Luce, 1982), results from the sum of independent Gaussian (Normal) and Exponential random variables. It has three parameters: Mu (μ), Sigma (σ) and Tau (τ), where μ and σ are the mean and standard deviation of the Normal component and τ is the mean (and standard deviation) of the Exponential component. These parameters have a simple relationship with the first three cumulants of the overall distribution: the mean of the ex-Gaussian equals the sum of μ and τ ; its variance is the sum of σ^2 and τ^2 ; and the third cumulant equals $2\tau^3$. Thus, the mean and variance of the overall distribution are a function of the location and spread of both the normal and the exponential components, while the exponential component alone gives rise to the asymmetry of the ex-Gaussian distribution (i.e., when $\tau=0$, the distribution is normal and skew increases with τ). Estimates of μ and τ therefore provide a concise summary of the location and skew of the overall distribution.

A second approach to measuring RT distribution shape involves estimating the distribution function locally. Again, Ratcliff (1979) has been influential in this approach with his adaptation of Vincent's (1912) technique (for related approaches see Silverman,

1986). For each subject, distribution shape is estimated by the mean RT in each of a series of equally probable ranges of the ordered sample of RTs. Each mean is called a vincentile. For example, 10 vincentiles can be obtained by averaging the first 10%, second 10% and so on of the ordered RT data. Vincentiles can be estimated using a simple algorithm given by Ratcliff (1979)³. Ratcliff (1979) argued that RT distributions should be estimated from vincentiles averaged over subjects when there are insufficient data points (less than 100 per subject) to support reliable maximum likelihood estimation for individual subjects. Average vincentiles approximately preserve the shape of underlying distributions for each subject, as long as the individual subject distributions are uni-modal and smooth (e.g., if each subject's RT distribution is ex-Gaussian, the distribution of their vincentile averages will be approximately ex-Gaussian). The parameters of an ex-Gaussian distribution fit to a vincentile average are also interpretable as the average of the corresponding parameters for each subject⁴. Vincent averages, therefore, provide a way of avoiding excessive variability in individual RT distribution shape estimates when experimental constraints, such as the number of words of a certain type, limit sample size. Plots of average vincentiles also provide a method of investigating mechanisms that influence particular local regions of the RT distribution. For example, a process that decreases mean RT by reducing the proportion of long RTs will influence the slower vincentiles but leave the faster ones relatively unaffected. We will use both average vincentile plots and ex-Gaussian parameters estimated from average vincentiles to provide converging evidence about RT distribution shape in different word identification tasks.

RT distributions and word identification

Two recent studies have applied RT distribution analysis to test more refined predictions about the processes underlying word identification than can be evaluated using the traditional measures of central tendency. Plourde and Besner (1997) used the LDT to investigate WFE and stimulus quality effects on RT mean and variance as well as on ex-Gaussian estimates of μ , σ and τ . They found that both μ and τ were larger for LF than HF words. Degraded compared to clearly presented items were also associated with higher μ and τ although the stimulus quality effect on τ was confined to participants with

a high error rate. Consistent with many investigations of the joint influence of these variables on mean RT (e.g., Becker & Killion, 1977; Borowsky & Besner, 1993), their effects on variance and the ex-Gaussian parameter estimates were additive rather than interactive. Using Sternberg's (1969) additive factors logic, Plourde and Besner argued that the lack of interactions between frequency and quality on any component of the RT distribution indicates that word frequency and stimulus quality influence separate processes⁵ and strengthens the evidence that "the locus of the word frequency effect in lexical decision should not be thought of as affecting the word detector level" (p. 189). However, their data do not clearly distinguish between alternative accounts of the basis of the WFE because they rely entirely on the LDT paradigm. If additive logic is accepted, their results suggest that frequency affects a processing stage that is insensitive to stimulus degradation, but a number of models of word identification can accommodate the additivity of frequency and stimulus quality effects without abandoning the assumption that WFE reflects lexical access (e.g., Becker, 1976; Besner & Johnson, 1987, Paap et al., 1982).

Balota and Spieler (1999) recently applied RT distribution analysis to test predictions derived from Balota and Chumbley's (1984) model of the LDT. The greater proportion of slow responses based on the additional second-stage required for LF words should result in a more skewed RT distribution, and therefore larger τ estimates, for LF than HF words. These τ effects should be specific to the LDT because they reflect decision processes specific to the binary classification requirements of this task. Balota and Spieler also compared the RT distributions for repeated and non-repeated words and nonwords. If repetition increases an item's FM value, it should increase the chance that a LF word will fall above the upper response criterion and allow a fast "yes" response, but decrease the probability of a fast "no" response to nonword stimuli. These effects should be reflected in reduced τ for repeated compared to non-repeated LF words, but decreased τ for repeated nonwords.

Balota and Spieler's results broadly supported these predictions. Consistent with Plourde and Besner's (1997) data, the 63 ms WFE on mean RT in the LDT was due to relatively equivalent changes in both μ (29 ms) and τ (35 ms), indicating that the

distributions for LF words were both shifted in location and more skewed than those for HF words. The WFE on mean RT for the same items in a word naming task was considerably smaller (18 ms) and the ex-Gaussian analysis showed that it was due entirely to changes in μ with no evidence of the τ effect observed in the LDT. The apparent specificity of the τ effect to the LDT is compatible with the predictions of the Balota and Chumbley (1984) model. The LDT repetition effects showed the typically greater repetition effect on mean RT for LF than HF words (e.g., Scarborough, Cortese & Scarborough, 1977). The ex-Gaussian parameters revealed that the repetition effect for HF words reflected a change in μ alone, while for LF words it was due to a reduction in both μ and τ . By contrast, nonword repetition was associated with an increase in both μ and τ . The repetition effect on naming was weaker than in the LDT, and repetition effects were additive with both frequency and lexicality. The task differences are again compatible with the claim that the larger frequency and repetition effects observed in the LDT data reflect task-specific processes.

Although the broad features of these results are compatible with Balota and Chumbley's (1984) conceptual description of the decision process underlying the LDT, Balota and Spieler (1999) were unable to explicitly simulate the empirical RT distributions obtained in the LDT using the original two-stage account. A number of models assuming different distributions of first and second-stage response times were compared but none could reproduce the complete pattern of effects of frequency and repetition on μ and τ . Balota and Spieler argued that their results showed a "dissociation of effects of variables on μ and τ [which] suggests that different processes may be influencing different parameters in the RT distributions" (1999, p. 43). They therefore implemented a model that assumes different effects of both frequency and repetition on first and second stage responses. The parameters that allowed the most effective simulation assumed additive effects of frequency and repetition on first stage responses, but interactive effects on the second stage. Balota and Spieler identify this modified model with the view that frequency and repetition influence two different processes, one that is fast and automatic and yields additive effects of frequency and repetition, and one that is slower, more attentionally demanding and shows larger repetition effects for LF

words. This distinction is consistent with the differential WFE obtained in masked and unmasked priming paradigms (Forster and Davis, 1984; Rajaram & Neely, 1992) and the advantage for LF words in recognition memory tasks (Glanzer & Bowles, 1976).

However, Balota and Spieler acknowledge that other accounts might be able to explain the results as well as their “hybrid two-stage model”.

Rationale for the present study

The finding of a WFE on skew that appears specific to the LDT is consistent with Balota and Spieler’s broad claim that different factors contribute to the WFE in LDT and naming tasks. But it does not provide clear support for a frequency-sensitive decision process that is unique to the LDT, because the data do not definitively establish whether the differential WFEs in LDT and naming tasks are due to decision processes that exaggerate the WFE in the LDT, or to sublexical processes that reduce the WFE in word naming tasks (Paap et al., 1987), or to some combination of the two. Distinguishing these alternatives requires a broader array of tasks than those compared by Balota and Spieler (1999). In addition to LDT and word naming tasks, the present experiment included two additional tasks selected to shed light on the locus of frequency effects on RT distribution.

The Animacy Semantic Categorization (SEMCAT) task required participants to classify HF and LF words as animate or inanimate. Although it is clear that some semantic categorization tasks can be strategically influenced, they are often regarded as the hallmark of lexical retrieval (Forster & Shen, 1996). The relevance of this task in the present context is that it does not allow responses to be based solely on item familiarity. If LDT frequency effects are partly due to familiarity based decision processes, they should be larger than those obtained in the SEMCAT task primarily because of a larger τ component. If both tasks involve a common frequency-sensitive process, they should show similar effects of frequency across the RT distributions for different conditions.

In the Lexically Contingent Naming (LEXNAM) task (often called Go No-go naming) participants were presented with both words and nonwords and told to read the item aloud only if it is a word. This task provides an estimate of lexically-based naming performance (Forster & Davis, 1991): participants must determine that the item is a word

before pronouncing it and the task therefore combines the demands of the LDT and naming tasks (Hino & Lupker, 1996, 1998). However, the crucial difference between the LEXNAM task and the LDT is that contingent naming requires a unique word-specific response so, in contrast to the LDT, responses cannot be based on familiarity alone. Contingent naming results also provide evidence about the contribution of sublexical processes to performance in the standard naming task. If the smaller WFE observed in naming tasks reflects the impact of nonlexical assembly processes on pronunciation of LF words, the LEXNAM task should produce a WFE equivalent to the LDT because contingent naming responses cannot be based solely on nonlexical assembly. Thus, the contingent naming task provides valuable information about the basis of the WFE in both LDT and naming tasks because it precludes both the fast familiarity-based responses presumed to be available for HF words in the LDT, and the nonlexical pronunciation responses that can sometimes be made to LF regular words before lexical retrieval has completed. If both sets of task-specific processes play a role, then the WFE in the LEXNAM task will be larger than in either LDT or standard naming. This general prediction has been confirmed in contingent naming data for both English and Japanese words (Hino & Lupker, 1996, 1998). Comparisons of frequency effects on RT distributions in LEXNAM with those for the standard LDT and word naming tasks will provide more refined evidence about the relationship between the processes involved in each task.

EXPERIMENT 1

The same set of items was presented to a large sample of subjects in five different tasks, three of which required a naming response and two a binary classification. The three different naming tasks all required the same pronunciation response, but were designed to vary the extent to which lexical retrieval was a prerequisite for pronunciation. In the Word Naming (WNAM) task, participants were presented with only word items and told to read them aloud as quickly as possible. The Word and Nonword Naming task (W/NWNAM) was identical except that words were mixed with nonwords and participants were told to read aloud all items. Because nonwords must, in theory, be named nonlexically (e.g., Monsell, Patterson, Graham, Hughes & Milroy, 1992), these

two variants of the standard naming tasks have been argued to provide a means of demonstrating the involvement of nonlexical naming procedures which are assumed to play a greater role when words are named in an environment of nonwords (but see Lupker, Brown & Colombo, 1997). The W/NWNAM task also serves as a control for the Lexically Contingent Naming (LEXNAM) task because the two tasks presented identical word and nonword stimuli but in LEXNAM participants were instructed to only read the item aloud if it was a word. The LEXNAM task therefore requires participants to establish that the item is a word before emitting a pronunciation response and cannot, therefore, rely entirely on nonlexical assembly.

Both the LDT and the SEMCAT task required a binary classification response. In the LDT this classification must be based on the lexicality of the stimulus. The categorization task required a binary animate/inanimate classification because it placed fewer restrictions on the stimuli that could be included and because it is less vulnerable to the category priming effects that are argued to contaminate Balota and Chumbley's (1984) version of the paradigm (Monsell et al., 1989). Both tasks seem to require lexical retrieval but only the LDT allows familiarity-based responses because familiarity does not provide a basis for the animacy judgments required in the SEMCAT task.

The SEMCAT task required selection of stimuli that could unambiguously be classified as animate or inanimate. Animacy was therefore included as an independent variable to allow its effects on RT distributions to be compared with the effects of frequency. Monsell et al.'s (1989) semantic categorization task, which required a similar classification of items as "person" or "thing", found that responses to "person" (i.e., animate) words were faster than those to "thing" (inanimate) words and that the latter also showed a smaller frequency effect. They could only attribute these differences to possible problems in item selection, but argued that the "form of the [frequency] effect" (p. 55) and its interaction with animacy in the categorization task were sufficiently similar to that in the LDT to support their claim that there is a "common basis for the effect of frequency on RT in the two tasks" (p. 56). The present research explores differences in the effects of both frequency and animacy on different tasks in more detail to provide further evidence regarding this claim. If the faster responses to animate words

are taken to reflect reliance on “semantic dimensions or features” such as “personhood” or “animacy” (Monsell et al., 1989), then animacy effects provide an index of the retrieval of semantic information which can be used to evaluate the relative contribution of such information to faster and slower responses in different tasks. For example, if semantic information is activated more slowly than the orthographic and phonological information stored in the lexical system, animacy effects on tasks that do not require retrieval of such information – all but the SEMCAT task in the present battery - should be more marked for slower responses.

Method

Participants. The participants were 91 students (31 male) from the University of New South Wales, with normal or corrected to normal vision who completed the experiment for course credit. The data for 16 subjects were eliminated because they had learned English only after starting primary school or because they made more than 15% errors in two or more of the five tasks. Two further subjects’ data were eliminated from some tasks because of equipment malfunctions. The final data set consisted of 74 sets of data for all five tasks, except the LEXNAM task, which had 75 subjects.

Stimuli. The word stimuli were selected to conform to a factorial manipulation of word frequency (High/Low) and animacy (Animate/Inanimate). HF words occurred at least 50 times/million in Kucera and Francis’s (1967) norms and LF words had frequencies of no more than 20/million. A set of 125 items was selected to fill each of the four cells of the factorial design. The complete set of 500 words was then divided into five lists each containing 25 words in each frequency/animacy combination. An attempt was made to match the words in each list on number of letters, number of syllables, and word frequency. The length range across the item set was from 3 to 10 letters (1 to 4 syllables). It turned out to be impossible to equate the letter length of the HF and LF words while maintaining equivalence on other factors: the mean length for the five lists of HF words ranged from 5.6 to 6.0 letters (Mean = 5.82), while those for the LF items ranged from 6.2 to 6.8 (Mean = 6.48). Although small, this difference is significant ($t(498) = 4.62$). The average word frequencies of the HF lists were between 167.7 and 176.8/million and those for the LF lists ranged from 8.5 to 8.9/million. Animacy

categorizations were made by one of the authors (S.A.) and confirmed through discussion with students and colleagues. Efforts were made to ensure that the two sets of items could not be systematically differentiated by the presence or absence of affixes associated with animacy. A complete word list is presented in Appendix A.

Three sets of 60 nonwords were constructed for use in the three tasks requiring nonword stimuli (LDT, W/NWNAM and LEXNAM). The nonwords were derived from a computer-generated list constructed to match the letter and bigram distribution of English words. Nonwords were orthographically and phonologically legal and selected to match the letter and syllabic length distribution of the word lists and to contain affixes similar to those represented in the word items (e.g., troan, ocarice, follour, manton, shestin, nimicater). The same set of nonwords was attached to each of the five word lists to be used for a particular task, but different nonword sets were used for different tasks. Thus, as was the case for words, participants did not experience repetitions of any nonwords across the five task conditions. In all tasks that included nonword items, the ratio of words to nonwords was 5:3. Each participant was presented with a different stimulus lists in each of the five tasks. Across participants, each list was assigned a roughly equivalent number of times to each task so that RT data were obtained for all 500 words in all five tasks.

Procedure. Participants were tested individually in sound-attenuated cubicles. To avoid response confusion, participants always completed the three naming tasks and the two binary classification tasks in sequence, with the two button-press tasks either preceded or followed by the three naming tasks. The order of SEMCAT and LDT was counterbalanced within the two button-press tasks, and six different orders of the three naming tasks were approximately equally represented. Prior to each task condition, participants were presented with 15 practice trials to accustom them to the new task requirements and two buffer items preceded each list of experimental items. In all tasks, subjects were instructed to make their response as quickly as possible while avoiding too high an error rate.

The DMASTR⁶ software package was used to control stimulus presentation, collect RTs, and measures of classification accuracy. For the naming tasks, the experimenter

monitored participants' performance and recorded pronunciation errors and trials on which the voice key failed to trigger or was triggered by extraneous noise, so that these trials could be removed from the RT analysis. Stimulus presentation procedures were identical in all tasks. Items were presented in lower case in the center of a computer screen. Each item remained on the screen for 1200 ms and responses were collected for 3 seconds from stimulus onset. The next trial was initiated 3 seconds after the offset of the previous item. Fixed, rather than self-paced presentation and a relatively slow presentation rate were employed to allow identical presentation procedures in all tasks and to ensure that participants did not feel differentially time pressured in the more difficult task conditions.

Results and Discussion

All analyses were restricted to the data obtained with word stimuli because exactly the same words were used in all tasks. Nonword performance was not analyzed both because nonword responses were only obtained in two tasks (the LDT and W/NWNAM) and because different nonword sets were used in each task in which they were included. The RT analyses yielded five different measures: mean and variance, and the ex-Gaussian parameters μ , σ , and τ . Means and variances were calculated directly from the data using the usual sample formulae. Ex-Gaussian parameter estimates were obtained from vincentile values averaged over subjects using maximum likelihood estimation procedures implemented with the RTSYS software (Heathcote, 1996).

We also examined accuracy in all tasks and conditions. There were three possible types of errors across the five tasks. In the tasks requiring a naming response, problems with the voice key (e.g., triggering by external sounds, responses too soft to be detected) caused 1.9% of trials to be discarded before analysis. The remaining two types of errors were misses (failing to respond within the time limit) and wrong responses. Misses were very rare in the W/NWNAM and WNAM tasks (0.08% and 0.15%) and so were not analyzed. A wrong response could either be due to a mispronunciation (naming tasks) or to pressing the wrong response button (LDT and SEMCAT).

Unless otherwise indicated, all results described as significant refer to a null hypothesis sampling distribution probability of less than .05.

Correlational analyses. Correlational analyses have provided a major source of evidence about the overlap between the processes involved in different tasks (e.g., Forster & Chambers, 1973). They also allow evaluation of the effects of attributes that were not explicitly controlled during item selection. Table 1 presents the correlations between mean RT to each item in the five different tasks and measures of word length (in letters), word frequency (WORDFREQ; Kucera & Francis, 1967) and its logarithm (LOGFREQ).

INSERT TABLE 1 ABOUT HERE

The correlations between mean item RT in the different tasks range from .3 to .7. The weakest relationship is between W/NWNAM and SEMCAT and the strongest between WNAM and LEXNAM. LEXNAM has the highest correlation with every other task except W/NWNAM implying that the processes underlying performance in the LEXNAM task have the greatest overlap with those required in all other tasks. The relationship to performance in both the LDT and the standard naming tasks is not surprising because, logically, the task combines the demands of lexical classification and word naming but the strong relationship between LEXNAM and SEMCAT performance is less intuitively obvious. The correlations between length and task performance confirm that length has a greater impact on naming performance than on binary classification tasks. Moreover, the fact that this correlation is lower for the LEXNAM task, which was intended to preclude nonlexical naming strategies, is consistent with Weekes' (1997) claim that length effects on naming performance are due, at least in part, to nonlexical procedures. Performance in all tasks is more highly correlated with log frequency than with a raw measure of number of occurrences per million, consistent with previous evidence that RT has a stronger linear relationship with log than absolute word frequency (e.g., Whaley, 1978).

INSERT TABLE 2 ABOUT HERE

To determine the unique contribution of the various stimulus attributes and to investigate whether the items differed in factors relevant to phonological production, simultaneous regression analyses were conducted on each task (see Table 2) using the predictors of log frequency and length as well as a set of 10 binary variables coding phonetic characteristics of the initial phoneme of each word and a rated measure of

regularity⁷. The phonetic variables were those employed in Treiman et al.'s (1995) and Balota and Spieler's (1997) analyses of large-scale naming datasets. The regression analyses demonstrated significant unique contributions of length to standard naming but not the LEXNAM or button-press tasks. The cluster of phonological variables predicted substantial proportions of variance in all three tasks requiring a pronunciation response but had no effect on button-press performance. Log frequency was a significant predictor of performance in all tasks. In the LDT it was the only significant predictor and accounted for almost all systematic variance. The pronunciation response tasks were also sensitive to response-specific factors that accounted, in the case of the standard naming tasks for more total variance than log frequency⁸. Frequency still accounted for a substantial portion of the systematic variance in naming times, but considerably less than for the other three tasks. Both the lower unique contribution of frequency to the two naming tasks, and the fact that the LDT shows the largest unique contribution of frequency are consistent with the Balota and Chumbley (1984) account.

Overview of Parameter Estimates. In each task, we used the maximum number of vincentiles possible given the size of the sample of correct responses. This was generally about half of the number of trials within a condition (25). Table 3 presents RT mean and variance estimates, ex-Gaussian parameter estimates, and percent misses and wrong responses for each condition of each task. The data points and standard error bars in Figure 1 are the RT for each vincentiles averaged over subjects in each condition of the five tasks. The lines in each graph join the estimated vincentiles for the fitted ex-Gaussian distributions⁹.

INSERT TABLE 3 ABOUT HERE

For the standard naming tasks (W/NWNAM and WNAM) the ex-Gaussian distribution provides a very accurate description: ex-Gaussian vincentiles differed from averaged vincentiles by much less than one standard error in all cases. For the remaining tasks the ex-Gaussian also provides a very good fit for all but the slowest vincentile. The misfit for the LDT and the LEXNAM task mainly occurred for LF words, with the fitted ex-Gaussian distribution overestimating the subject average value. Overestimation of the

slowest vincentile by the ex-Gaussian distribution is greatest in the SEMCAT task, being in excess of one standard error for all conditions¹⁰.

INSERT FIGURE 1 ABOUT HERE

The misfit of the slowest vincentile in the SEMCAT task might be caused by missed responses due to the time limit on responding. The time limit can remove genuine slow responses to difficult words and has its greatest effect in SEMCAT because it is the slowest task overall, especially for LF words. Miss rates are also quite high for LF words in the LEXNAM task, but many of these misses are likely to be genuine error responses, that is, items for which subjects deliberately withheld a response because they thought that the item was not a word. To check the possibility that misfit was caused by missed responses in SEMCAT, we fit the ex-Gaussian distribution to vincentiles averaged over subjects who had fewer than 3% misses ($n = 55$)¹¹. As shown in Figure 1f, the fit of the ex-Gaussian was improved and is similar to the fit for LEXNAM and LDT. Analyses of the ex-Gaussian parameters for the fits to low miss subjects in the SEMCAT task produced the same pattern of significance as analyses of parameters from fits to all subjects. Because the pattern did not differ, and in order to maintain comparability between subject samples for each task, subsequent data and analysis are based on averages over all subjects in the SEMCAT task.

To confirm and extend analyses of the ex-Gaussian, we analyzed individual subject vincentile values. These analyses used 10% vincentiles for all tasks so that the fastest and slowest vincentiles could be compared across tasks. In all inferential tests, task was treated as a between-subjects factor whereas frequency and animacy were treated as within subjects-factors¹². Inferential tests of the ex-Gaussian parameters were constructed by resampling (Efron & Tibshirani, 1996) using a modified version of the RTSYS software¹³.

Inter-relationships between measures. The overall results and relationships between the various measures are first presented descriptively to provide an overview of the results for each task and to demonstrate that the different measures, while often correlated, yield unique information. Figure 2 plots RT variance against mean RT (Panel A), τ against μ (Panel B), and total error rate against mean RT (misses + wrong

responses) (Panel C). Within each task, a line joins HF and LF conditions of the same type (solid lines for animate words and dotted lines for inanimate words). Hence, the length of each line indicates the combined magnitude of the WFE across both measures, while the distance between the lower (HF) and upper (LF) end of each line, on either axis, indicates the magnitude of the WFE on that specific dependent variable. Within each task, the WFE across measures is strongly positively correlated, except for errors in SEMCAT. For all but inanimate words in SEMCAT, the WFE is large and positive (i.e., all measures are higher for LF than HF items). Animacy effects, which can be seen in Figure 2 by joining the corresponding ends of the solid and dotted lines for each task, are weaker than frequency effects in all tasks except SEMCAT. The direction of all animacy effects is positive (i. e., inanimate larger than animate), except for errors to LF words in SEMCAT. SEMCAT also differs from the other tasks in the relationship between mean RT, variability and skew. RT standard deviation and τ increase at about half the rate of mean RT for all tasks except SEMCAT, where the rate was about one third to one quarter. Perhaps the most surprising feature of the summary data is that the overall central tendency (mean RT and μ) of the RT distributions is faster for LDT than naming. Explanations of this discrepancy with the usual finding that LDT responses are slower than naming latencies (e.g., Forster & Chambers, 1973) are considered in the General Discussion.

INSERT FIGURE 2 HERE

Comparisons between tasks show several unique effects on the different measures. Overall for each task, location measures are smallest for the LDT, second for the W/NWNAM and WNAM tasks, third for the LEXNAM task, and largest for the SEMCAT task. In contrast, variance, error rates¹⁴ and τ are smallest for the WNAM and W/NWNAM tasks, second - and approximately equivalent - for the LEXNAM and LDT tasks, and largest for the SEMCAT task. Thus, variability and skew are not simple correlates of mean RT. The inter-relationships between the various measures suggest three task clusters. The W/NWNAM and WNAM tasks are almost identical on all measures. The measures that summarize the location of the RT distribution - mean RT and μ - differentiate LEXNAM and LDT. LDT responses are much faster, but these two

tasks are almost identical on all other measures. Performance in the SEMCAT task stands alone.

The following three sections describe the results of statistical analyses of the various dependent measures separately for the three task clusters. The dependent measures differ in precision so tests of the same effect will not have the same power for different measures. We provide effect magnitudes as well as the results of inferential tests on the effects so their sizes can be compared without confounding by differences in power. In each analysis, the magnitude of the main effects and interactions were defined such that the most common outcome yielded a positive value (i.e., a positive value for the frequency and animacy effects indicates slower responses to LF and inanimate words, respectively, and a positive interaction indicates that the animacy effect was larger for LF words).

W/NWNAM and WNAM tasks. Table 4 presents the grand means and the magnitudes of each main effect (task, frequency and animacy) and their interactions, along with the results of inferential testing of the significance of the effects observed in each task. Errors consisted only of mispronunciations. No effect involving the task factor was significant indicating that the presence of nonwords had minimal effect on the way subjects named words (Keefe & Neely, 1990). The WFE was large and highly significant for all measures except σ . These results contrast with Balota and Spieler's (1999). They found that the WFE on mean naming RT was less than half as large and occurred only in μ , whereas we observed a highly significant WFE on τ . The fact that frequency exerted a highly significant effect on variance but not on σ demonstrates that the variance effect is primarily due to the increased τ for LF words, rather than to an overall increase in performance variability. LF words were also much more likely to be mispronounced than HF words.

INSERT TABLE 4 ABOUT HERE

The main effect of animacy and the interaction of animacy with frequency are quite small but highly significant in mean RT. Figure 1 shows that both the main effect and the interaction are confined to LF words and slower vincentiles. To further investigate this observation, we performed repeated measures ANOVAs on 10% vincentile values

calculated for each subject. Figure 3 plots the magnitude of the frequency and animacy main effects and of the interaction between frequency and animacy across the ten vincentiles for each task. A flat line indicates an effect on only the location (mean and μ) of the RT distribution, whereas a change with vincentile number indicates an effect on variance or skew. The increase in frequency and animacy effects across vincentiles are consistent with the main effects of frequency and animacy on variance, and the main effect of frequency on τ in Table 5. Figure 3 also confirms the impression from Figure 1 that animacy only affects LF words, and that this effect increases for slower responses. We also calculated three-way ANOVAs using the first and last vincentile values as dependent variables to examine effects on very fast and slow responses. There were no differences between the W/NAM and W/NWNAM RT distributions. The frequency main effect on the first vincentile was highly significant, indicating that frequency influenced even the fastest naming responses but neither the animacy effect nor the frequency by animacy interaction were significant for the first vincentile, confirming that animacy did not affect fast responses for either LF or HF words. For the last vincentile, in contrast, the main effects of frequency and animacy, as well as their interaction, all yielded significant effects.

INSERT FIGURE 3 ABOUT HERE

In summary, inferential analyses of the W/NWNAM and W/NAM tasks showed that none of the measures significantly differentiated the naming tasks, and no effect or combination of effects interacted with task on any measure. Frequency had a large effect on all measures except σ . Frequency differences are evident from the first vincentile and increase with later vincentiles for both animate and inanimate words. Animacy does not affect HF words in either task. For LF words, animacy does not affect fast responses, but it has an increasing effect as responses become slower.

LDT and LEXNAM tasks. Table 5 presents the grand means and the magnitudes of each main effect (task, frequency and animacy) and their interaction, along with the results of inferential testing. Figure 4 plots the frequency and animacy main effects and their interaction across the ten vincentiles for each task. Task had a highly significant

effect on the location measures, mean RT and μ , but did not approach significance for either variance or τ . Analyses of the first and last vincentiles also showed highly significant main effects of task but no interactions with task approached significance, except for a three-way interaction between task, frequency and animacy for the first vincentile. As can be seen in Figures 1, the interaction occurs because animacy affects only LF words in the LEXNAM task, but has a slightly larger effect on high than LF words in the LDT. Apart from this difference, the results show that the RT distribution for the LEXNAM task is shifted to be approximately 125 ms slower than that for the LDT, with virtually no change in distributional shape.

INSERT FIGURE 4 ABOUT HERE

The error analysis in Table 5 compared the sum of misses and mispronunciations in the LEXNAM task with the sum of misses and wrong responses in the LDT. Average error rate was 1.2% higher in the LDT than in the LEXNAM task, but the effect was only marginally significant. Misses were more common in LEXNAM (3.54%) than in the LDT (0.15%), and wrong responses in the LDT (5.28%) were more common than mispronunciations in the LEXNAM task (0.72%). We compared summed error measures because the response rule in the LEXNAM task (i.e., respond only if the item is a word) means that misses could be due to either wrong or slow decisions.

INSERT TABLE 5 HERE

The WFE was highly significant for all measures except σ . Replicating Balota and Spieler (1999) and Plourde and Besner's (1997) LDT findings, the main effect of frequency was greater for τ than μ , and small for σ . In fact, our LDT ex-Gaussian estimates are very similar to those in the earlier reports: we found μ , σ and τ frequency effects of 36, 6 and 46 msec, respectively, compared to values of 20, 3, 37 (Balota & Spieler, 1999) and 29, 6, 35 msec (Plourde & Besner, 1997). The animacy main effect was smaller but still significant on mean and variance, and also on τ . In both tasks, interactions between frequency and animacy reflected a greater animacy effect for LF words.

Figure 4 demonstrates that animacy and frequency effects increased with RT, and that the animacy effect tended to increase more for LF than HF words. Separate analyses

of the first and last vincentile showed highly significant main effects of frequency and animacy on both fast and slow responses, whereas the interaction of frequency and animacy was significant only in the last vincentile.

The SEMCAT task

The effects of frequency and animacy on SEMCAT performance, and particularly their interaction, are in marked contrast to the other tasks. Table 7 reports effect magnitudes and significance test results for the various measures, and Figure 5 plots the effects as a function of vincentile number. Frequency had significant effects on all measures including σ . The WFE on mean RT and errors was similar in magnitude to the LDT and LEXNAM tasks and larger than that observed in the W/NWNAM and WNAM tasks. However, the difference between mean RT for HF and LF words reflected different distributional properties from those underlying the lexical discrimination tasks. The WFE on μ was larger than in any other task, whereas for τ it was almost identical to the small effect found in the W/NWNAM and WNAM tasks. The main effect of animacy on location measures, mean RT and μ , was also considerably larger than the other tasks but, τ was unaffected by animacy. This contrasts with the other tasks where the animacy main effect was mainly due to increases in τ for inanimate words. Responses to inanimate words were also more variable than responses to animate words in all tasks. In the SEMCAT task this was due to an increase in σ whereas in the other tasks it reflected an increase in τ . Thus, unlike the lexical discrimination tasks, the slower classifications of inanimate words in SEMCAT are not confined to slower RTs. This is shown in the flatter slopes of the frequency and animacy functions in Figure 5 compared to Figure 4.

INSERT TABLE 6 ABOUT HERE

The interaction between frequency and animacy in mean RT was small, but highly significant. It was in the opposite direction to all other tasks, and so is negative in Figure 5. Specifically, the advantage for animate over inanimate items was larger for HF than LF words, mainly because of a marginally significant effect on μ . Analyses of the first and last vincentiles showed that animacy and frequency had highly significant effects on both fast and slow responses. The reversed interaction between frequency and animacy was significant for the first vincentile, and reflected a 25 ms larger animacy effect for HF

than LF words. The effect for the last vincentile was of the same magnitude (28.5 ms) but not significant implying greater variability in the last vincentile. Inspection of Figure 5 confirms that, like the main effect of animacy, the reversed interaction was due to a shift in the overall location of the distributions rather than being confined to slow responses and therefore does not appear to be an artifact of the relatively poorer ex-Gaussian fit for the slower vincentiles.

INSERT FIGURE 5 ABOUT HERE

A possible explanation of the reversal of the interaction relative to the other tasks is suggested by the pattern of error rates. This measure showed the strongest evidence of the reversed interaction. In contrast to all the other tasks, HF words show a larger animacy effect on total (misclassifications + misses) error rate (Animate = 5.5%, Inanimate = 11.0%) than LF words (Animate = 13.7%; Inanimate = 9.9%). The error rate for HF inanimate items is particularly striking because the same items yielded less than 2% errors in all other tasks. LF words also showed a reversal of the usual trend towards more accurate performance for animate than inanimate items. This apparently reflected a speed-accuracy trade-off because, as can be seen in Figure 2c, LF animate words were classified more quickly, but less accurately than LF inanimate words.

The pattern of speed-accuracy relationships in SEMCAT suggests that participants were biased to respond “yes” (i.e. animate) to familiar (usually HF) words and “no” (inanimate) to less familiar (usually LF) words. A tendency to respond “inanimate” to LF words explains the high error rate for LF animate items as well as the reduced frequency effect on error rate for inanimate words: participants are biased to make fast, but incorrect, “animate” responses to HF words while the bias to respond “inanimate” to LF words enhances accuracy for this condition and reduces the WFE. Consistent with this interpretation, the misclassification rate in the LF inanimate condition was lower than in the LDT (6.4% vs. 9.4%) even though SEMCAT yielded a higher error rate than the LDT in every other condition. Logically, the misclassification rate should be higher for semantic categorization than lexical classification because participants cannot make semantic judgments about items that they do not know to be words. The fact that this does not hold for LF inanimate words is consistent with the view that low familiarity

leads to a bias to respond “inanimate”. The decision bias can also account for the reversed interaction effect on RT because the two conditions in which the familiarity bias could enhance speed of classification are HF animate and LF inanimate words. The decision bias would therefore increase the RT animacy effect for HF words and reduce it for LF words giving rise to the observed interaction of a larger animacy effect for HF than LF words.

Further insight into the nature of the familiarity-bias derives from comparisons of RT for correct and incorrect responses. Incorrect responses to inanimate words of both HF and LF were faster (47 ms and 28 ms, respectively) than those for corresponding correct responses. In virtually every other condition¹⁵ of every task, wrong responses and mispronunciations were *slower* than corresponding correct responses (Mean difference: 167 ms, Range:41-366 ms). Thus, the “animate” responses that subjects erroneously made to inanimate words were faster than their correct responses to these items, while incorrect “no” responses to animate words were 81 ms slower than correct responses to the same items. This pattern of relationships is consistent with the possibility that “animate” responses sometimes reflect a fast familiarity-based process while judgments that a word is not animate rely on the outcome of a different, more slowly operating process.

The hypothesized tendency to make responses on the basis of item familiarity could be specific to conditions in which participants performed a series of tasks with different response requirements. In particular, half of the participants completed the SEMCAT task immediately after the LDT so the familiarity bias could reflect some form of carry-over effect. To test this possibility, the pattern of mean RT data was compared for participants who completed SEMCAT before (N= 36) or after (N=38) the LDT. The only significant difference between the groups was a marginal interaction between task order and the animacy effect for HF words ($F(1,72)= 4.09$) reflecting a somewhat larger animacy effect on RT when SEMCAT followed the LDT. This is consistent with an enhanced familiarity bias, but there were no accompanying changes in error rate, and the overall patterns of both speed and accuracy were identical regardless of order. Thus, while the familiarity

bias might be enhanced by prior exposure to the LDT, it does not appear to be solely due to carry-over effects.

Differences between task clusters. To provide further insight into the nature of the differences between the three task clusters, ANOVAs were conducted comparing overall performance, and the effects of frequency, animacy and their interaction on each dependent measure for each pair of tasks. The results of these comparisons are presented in Table 7. The W/NWNAM and WNAM tasks did not differ significantly on any measure. The only significant differences between LDT and LEXNAM were in the overall location of the distributions – LEXNAM responses were about 125 ms slower than lexical classifications - and the slightly larger frequency effect on errors observed in the LDT.

INSERT TABLE 7 ABOUT HERE

The naming and lexical discrimination task clusters differed in overall location and distributional shape, but nevertheless yielded similar patterns of effects across the various measures. The overall RT distributions for the LDT and LEXNAM tasks were more skewed and more variable than those for the two naming tasks, even though the LDT was significantly faster than the naming tasks in mean RT and μ , while LEXNAM was significantly slower on these location measures. The other major difference between the two task clusters is that frequency, and to a lesser extent, animacy had stronger effects on LDT and LEXNAM than on the standard naming tasks. The large difference in the effect of frequency on mean RT that was a focus of this research is due to changes in both the location and the shape of the RT distributions for HF and LF words. The discrimination tasks showed larger frequency effects on μ and τ than either naming task and, although the WFE was greater on μ than τ for the W/NWNAM and WNAM tasks, it was greater on τ than μ in the LDT and LEXNAM tasks. This replicates Balota and Spieler's (1999) finding of a larger τ frequency effect in LDT than naming.

Despite the differences in the general RT distributions associated with naming and discrimination tasks, they also show similarities. Frequency effects were larger than animacy effects and both were larger in later vincentiles and were therefore reflected in τ as well as in location measures. The tasks in both clusters also showed an interaction

between frequency and animacy due to a larger animacy effect for LF words, which manifested in all tasks primarily in measures of location rather than skew. This common pattern of frequency and animacy effects is different from that observed in SEMCAT.

SEMCAT mean RT was substantially longer than for any other task but again this difference was due to different changes in distribution shape for different task comparisons. The overall SEMCAT distribution was higher in the location (μ) and variability (σ) of the body of the distribution, and was more skewed (τ), than the two naming tasks, with the biggest difference in τ . SEMCAT was also significantly different from the LDT on all three ex-Gaussian parameters, but the largest difference was in μ rather than τ . The only ex-Gaussian parameter that significantly differentiated SEMCAT from LEXNAM was τ : the overall skew of the SEMCAT distribution was greater than for LEXNAM. In addition to these overall differences in distribution shape, SEMCAT also differed from each of the other tasks in the effects of frequency, animacy and their interaction on mean RT. For the two standard naming tasks, the differences between all of these effects were due to changes in μ . Location also accounted for the differential animacy and interaction effect when SEMCAT was compared with LDT and LEXNAM, but the differential WFE in these tasks was primarily due to the larger τ component to the frequency effect for the two tasks requiring lexical discrimination. SEMCAT also differed from the other tasks in the direction of the frequency by animacy interaction on RT and in the presence of an interaction in the pattern of errors, which was not evident in any other task. These last differences appear to reflect some form of familiarity bias as described above, but such a bias does not provide an explanation of the greater overall skew of the SEMCAT data.

EXPERIMENT 2

Three of the tasks in the Experiment 1 battery relied on pronunciation responses that were measured using vocalization onset latency recorded by a voice key triggered by an above-threshold vocal input. As demonstrated in the regression analyses of Experiment 1, such measures are influenced by the phonological characteristics of the initial phonemes of target items (e.g., Treiman et al., 1995). The differences between tasks could, therefore, be due to pronunciation-specific factors. The task most frequently used to

estimate the contribution of such factors is the “delayed naming task”, in which participants are presented with a stimulus word but told to withhold pronunciation of the item until they are presented with a cue. With sufficiently long delays between stimulus presentation and the cue to the pronounce the word, pronunciation latencies in this task should be influenced only by variables related to production.

Method

Participants The participants were 24 different individuals from the same population as tested in Experiment 1.

Stimuli and procedure Only word stimuli were presented in this experiment. The stimuli were the five word lists from Experiment 1. Items were displayed in the center of the screen for 1200 ms exactly as in the naming task of Experiment 1. Participants were told not to read the word when it was presented, but to delay producing their pronunciation until the appearance of a pronunciation cue (***) . The cue was presented following stimulus offset in the same location as the target item. To ensure that participants could not anticipate the cue, the delay between target offset and the pronunciation cue randomly varied among 3 intervals of 600 ms, 900 ms, 1200 ms. Five different lists that varied the assignment of items to target-cue intervals were presented to approximately equal numbers of participants.

Results and Discussion

Trials spoiled by external noise or voice key failures (0.88%) and anticipatory responses with RTs of less than 200 ms (1.25%) were removed before analysis, and RTs for error responses were removed from the latency analysis. RT mean and variance and estimates of the ex-Gaussian parameters were obtained and analyzed using the same procedures as for Experiment 1. Table 8 presents the average estimates of the various dependent variables for each condition and Figure 6 summarizes the vincentile averages and the fitted ex-Gaussian distributions. The fit of the ex-Gaussian was excellent, being well within one standard error of the average vincentile value for each vincentile.

INSERT TABLES 8 AND 9 ABOUT HERE

Table 9 reports effect magnitudes and significance test results for the various measures. The only significant effect of frequency is on total errors. As in Experiment 1,

more errors were made to LF words. Two effects were significant in the response time measures: mean RT and τ were 9ms and 16ms less, respectively, for inanimate than animate words. This contrasts with the animacy effects observed in the standard naming tasks where faster responses occurred for animate than inanimate words. Examination of Figure 6 confirms that the effect was confined to the slower vincentiles.

Analyses comparing the delayed naming data with those for the three pronunciation tasks of Experiment 1 (WNAM, W/NWNAM and LEXNAM) showed that the delayed naming distribution had a faster leading edge (μ) and a smaller μ frequency effect than any of the other tasks. The delayed naming distribution was also significantly less skewed overall, and yielded a smaller τ frequency effect than either WNAM or LEXNAM. Reflecting the opposite effects of animacy, delayed naming also differed significantly from the other three tasks in the animacy effect on τ .

Thus, in contrast to some previous applications of the delayed naming paradigm (e.g., Balota & Chumbley, 1984; Balota & Shields, 1990; Goldinger et al., 1997), frequency did not influence delayed naming performance. Given that the delay between stimulus presentation and the pronunciation cue was at least 1800 ms, this is consistent with Monsell et al.'s (1989) claim that frequency does not influence delayed naming performance when there is sufficient opportunity for articulatory preparation. The absence of a WFE on delayed naming RT also indicates that the WFEs observed in the other pronunciation tasks are not due to production-related characteristics of the initial phonemes of HF and LF words. There does appear to be a systematic production-related difference between animate and inanimate words reflected in the significant animacy effects on mean RT and τ . However, this differences appear to benefit inanimate over animate words and cannot, therefore, account for the significant animacy advantage observed in the other pronunciation tasks. Rather, the production differences may have led to an underestimate of impact of animacy on naming performance.

General Discussion

This is the first time that performance for the same set of items has been compared across such a broad array of word identification paradigms. Before discussing the additional insights provided by the RT distributions, it is appropriate to evaluate the mean

RT data that have provided the basis of most previous cross-task comparisons of the WFE. Paralleling previous research, the WFEs in both WNAM and W/NWNAM (51 and 46 ms, respectively) were smaller than in the LDT (76 ms). The SEMCAT task yielded a WFE on mean RT (70 ms) that was statistically equivalent to the LDT. This result is consistent with previous research using a semantic categorization task requiring animate/inanimate classifications (Monsell et al., 1989; Forster & Shen, 1996) but not with some other semantic judgment tasks (e.g., Balota and Chumbley, 1990). The most novel addition to this task comparison is the LEXNAM task, which also produced a WFE equivalent to the LDT (86 ms).

Within a conventional view of the lexical processing architecture (Monsell et al., 1989), the mean RT data imply that the WFE in the LDT reflects a lexical process that is also required for both LEXNAM and SEMCAT. This implies that standard naming tasks underestimate the lexical WFE because of the contribution of nonlexical pronunciation assembly procedures (Paap et al., 1987). Participants must avoid relying on the nonlexical procedure in the LEXNAM task so it produces a lexical WFE equivalent to the LDT. The correlational data support the view that naming is influenced by a serial pronunciation assembly process (Rastle & Coltheart, 1999; Weekes, 1997) that does not contribute to pronunciation responses in LEXNAM because word length was significantly correlated with performance in standard naming tasks but not LEXNAM. However, the lower unique contribution of frequency to the regression equation for SEMCAT compared to the LDT (12.9% vs 27.1%) conflicts with the mean RT data showing statistically equivalent WFEs and contradicts the conventional view that SEMCAT and LDT performance reflect a common lexical access process. This discrepancy reflects the differential impact of word frequency on the RT distributions for SEMCAT by comparison with the other tasks which is discussed in detail below.

One feature of the mean RT data that conflicts with previous research (e.g., Forster & Chambers, 1973) is the finding that LDT classification time (566 ms) was faster than that for the naming tasks (WNAM= 595 ms; W/NWNAM= 606 ms). We cannot conclusively determine why LDT responses were faster than word naming, but it might reflect the use of a stimulus set consisting of relatively long, primarily regular words and

matched nonwords. This contrasts with much of the existing research comparing LDT and naming tasks which has relied on monosyllabic words and nonwords deliberately selected to over-represent irregular and/or inconsistent words. There are two reasons why this stimulus set might eliminate the speed advantage for naming over LDT responses. First, the relatively small proportion of ambiguous words and nonwords, may encourage adoption of a less cautious lexical decision criterion than applied in many previous experiments and hence reduce overall RT (Stone & Van Orden, 1993). This may have been enhanced by a bias to respond “word” encouraged by the use of a word:nonword ratio of 5:3. Secondly, the use of relatively long stimuli might differentially slow naming compared to LDT responses. Both LDT and naming responses are slower for longer items (Forster & Chambers, 1973), but length effects are more marked in naming than the LDT (Fredriksen & Kroll, 1976). The regression analyses in Table 2 confirm the significant contribution of length to WNAM and W/NWNAM latencies and show no effect of length on LDT.

The apparent discrepancy with previous findings is unfortunate because it raises questions about the comparability of the present data to previous literature. However, it is also an advantage in the present context because it rules out the possibility that the larger WFE observed in the LDT is simply a function of longer average RT: the WFE is larger in LDT than naming tasks even though overall LDT classification speed is faster than naming time. With this single exception, the mean RT data for the five tasks are compatible with previous research on cross-task differences in the WFE. However, this measure may obscure differences between the RT distributions for the different tasks.

The data for all tasks were very well fit by the theoretical ex-Gaussian distributions estimated using maximum likelihood procedures. The ex-Gaussian estimates were supplemented by analyses of mean RT data across vincentiles that are independent of ex-Gaussian assumptions. These were completely consistent with the results of the ex-Gaussian analyses and provide a more direct way of relating distributional changes to the predictions of different models.

Every task in the present battery yielded slower and more skewed distributions for LF than HF words. The skew component of the WFE was larger in the LDT than the

naming task, however, in contrast to Balota and Spieler's (1999) finding that the naming WFE was entirely due to μ , the present WNAM and W/NWNAM data revealed a significant WFE on both μ and τ . This may reflect the use of items that were longer, on average than Balota and Spieler's (1999).

The results from this task battery provide evidence relevant to four questions that are crucial to distinguishing different interpretations of cross-task differences in the WFE. First, comparisons of the LDT with the other tasks provide evidence about the contribution of task-specific lexical decision processes. Second, comparing the three tasks requiring pronunciation responses with the two binary classification tasks sheds light on the role of nonlexical processes in naming performance. Third, comparing the results for the SEMCAT task with the remainder of the battery provide insight into how semantic information contributes to word identification performance. Finally, we address to the general question of how effects of familiarity relate to "normal" word identification processes.

Does LDT performance reflect decision processes rather than lexical access?

The LDT data replicate the increased skew for LF words reported by Plourde and Besner (1997) and Balota and Spieler (1999), but this component of the WFE is not specific to the LDT. The LEXNAM task produced a WFE almost identical to the LDT despite being shifted in location by about 125 ms. The RT distributions for LF words in the SEMCAT task were even more skewed than for the LDT but, because the SEMCAT distributions for HF words were also heavily skewed, the WFE in this task manifested primarily in μ .

Balota and Spieler (1999) acknowledge that many frameworks predict that RT distributions will be more skewed for LF than HF words. In some models skew effects arise from general word identification processes rather than from LDT-specific mechanisms. For example, parallel activation frameworks that assume an exponential activation function (e.g., Morton, 1970) predict a larger proportion of slow responses for LF items with higher thresholds (Balota & Spieler, 1999). Similarly, the competitive lateral inhibition process assumed by IA frameworks like MROM implies that inhibition from similar words will be greater for LF words. LF words should therefore yield a more

skewed RT distribution than HF words. Grainger and Jacobs (1996) found that the distribution of response times simulated by MROM underestimated the skew of the empirical data, for both correct and incorrect word responses. Because their experiments and simulations used only LF words, it is unclear whether this reflects a general underestimation of the proportion of slow lexical classification responses, or a specific underestimate of the enhanced skew for LF words that is observed in the LDT.

Frequency effects on “1st stage” LDT responses. Balota and Spieler’s (1999) evaluation of the LDT concentrates on the implications of the differential skew predicted by the 2nd stage of the original Balota and Chumbley (1984) model. Little attention is paid to seeking explicit evidence of the fast first stage responses that are predicted to arise from the familiarity mechanism of Balota & Chumbley (1984) and the overall activity mechanisms of MROM and DRC.

All of these models of the LDT task predict a reduced WFE on the leading edge of the distribution of LDT responses because all items that elicit FM values (Balota and Chumbley, 1984) or summed lexical activity measures (Grainger & Jacobs, 1996) that exceed the relevant criterion will yield fast lexical classifications, regardless of their frequency (Gordon, 1985). These fast familiarity-based word classifications mean that the WFE in the fast tail of the LDT distribution should be smaller than for the LEXNAM and SEMCAT tasks because the latter responses cannot be derived from familiarity alone. Even though the LEXNAM task is contingent on word/nonword discrimination, the unique pronunciation response required cannot be determined from familiarity. Similarly, the appropriate semantic categorization of an item cannot be determined from high familiarity because, unlike the LDT, there is no correlation between familiarity and the required animacy judgment.

The vincentile data presented in Figures 1, 4 and 5 clearly show that, although the leading edge of the LDT distribution is faster than for the LEXNAM and SEMCAT tasks, there is no evidence of the predicted difference in the magnitude of the WFE on fast responses. All three tasks showed a significant WFE on the first vincentile and, although the magnitude of the effect was slightly smaller in the LDT (33 ms) than in either SEMCAT (39 ms) or the LEXNAM task (43 ms), none of the differences between tasks

was significant. Thus, at least in terms of this operationalization of how the LDT-specific familiarity process should be reflected in the RT data, the results provide no evidence of such a process, or at least no evidence that it is unique to the LDT. The greater skew in the slow tail of the RT distribution for LF words in both the LDT and LEXNAM tasks is compatible with a task-specific decision process like that described by Balota and Chumbley (1984), but it could also arise from general word identification processes that are common to a variety of tasks (Paap et al., 1987).

Is the reduced WFE in naming tasks due to nonlexical processes?

Word naming responses are item-specific and cannot, therefore, be based on familiarity or overall lexical activation alone. However, as dual route frameworks emphasize, they do not necessarily require lexical retrieval. Although the precise predictions of dual route models depend on specific assumptions about the timecourse and interaction between routes (Stone & Van Orden, 1993), all variants share two fundamental assumptions that yield predictions about the relationship between the RT distributions for naming and LDT tasks. First, nonlexical processes operate more slowly, on average, than lexical processes; and secondly, word frequency only influences the lexical procedure. The nonlexical route therefore has more influence on pronunciations of LF words, accounting for the robust finding that phonological influences on pronunciation - as reflected in effects of variables like regularity and consistency - are more marked for LF words (e.g., Andrews, 1982).

When extended to RT distributions, these assumptions predict that nonlexical influences should lead to truncation of the distribution of RTs by comparison with tasks that must rely on the lexical procedure. Moreover, the truncation will be primarily due to reduced skew for LF words because assembly-based pronunciations will reduce the proportion of responses that must await the outcome of the slow lexical retrieval that is characteristic of LF words. When applied to the differential WFE in LDT and naming tasks, this is identical to the Balota and Chumbley (1984) account because it predicts a reduced skew component to the WFE in tasks that allow non-lexical responses. However, the dual route model also yields a prediction about in naming performance that is not confirmed by the present data. If naming responses to LF words are sometimes based on

nonlexical assembly rather than lexical retrieval, then the WFE within the naming task should reduce across the RT distribution. The longer it takes for lexical retrieval to complete, the greater the chance that nonlexical assembly will complete before retrieval so the WFE reduction should be most marked in the slow tail of the RT distribution. Figure 3 shows that this is not the case: the WFE in both WNAM and W/NWNAM increases in slower vincentiles, but not as much as in the tasks requiring lexical discrimination.

However, this feature of the naming data cannot be taken as strong evidence against dual route models because the prediction that the WFE will reduce for slower responses only holds for regular words. For irregular words, lexical and nonlexical processes yield different outcomes. Such conflicts will be more common for LF words and might therefore enhance skew for irregular words and counteract the reduction in the WFE predicted for regular words. Most items in the present sample were regular, but the presence of irregular items undermines a strong prediction that the WFE should reduce for slow naming responses. Another stimulus characteristic that might explain the increased WFE on slow naming responses is word length. Nonlexical assembly processes operate serially and yield longer completion times for longer words (Rastle & Coltheart, 1999). For HF words, the lexical route usually completes first so length will have little impact on performance. For short LF words, the nonlexical route may complete before the lexical route and contribute to reducing the WFE, but long LF words will be dealt with slowly by both routes. Thus, because the present sample included relatively long words, the nonlexical process may rarely complete substantially more quickly than the lexical route even for LF words. This could explain why there is no evidence of the reduced WFE for slow naming responses that dual route models predict for short regular words. However, this qualification would undermine the dual route account of the difference between LDT and naming tasks because assuming that naming responses for long LF words are lexically mediated implies that the two tasks should yield similar WFEs.

The serial nature of the nonlexical process in dual route models may also contribute to explaining task differences in the WFE on the fast tail of the RT distributions. The two

naming tasks showed a significantly smaller frequency effect on the first vincentile (W/NWNAM= 21 ms; WNAM= 22 ms) than any of the other tasks (33-43 ms). This appears consistent with Kawamoto and Zemblidge's (1992) proposal that the nonlexical procedure yields a distribution of finishing times with a slower mean, but a larger variance than the lexical route. Monsell et al. (1992) suggested that the differential variance might reflect a nonlexical assembly process, which "works through the input chunk by chunk" (p. 463) so that "delivery of the assembly process's opinions about segmental phonology [is]... more spread out over time, reflecting its essentially left-to-right operation" (1992, p. 464). The implications of this seriality assumption depend on whether non-lexical pronunciation is initiated as soon as the beginning segment(s) of the string achieve threshold or only once a pronunciation for the complete string is available. The faster leading edge of the distribution of nonlexical completion times that is implied by the reduced WFE on fast naming responses might reflect fast articulation of initial phoneme segments activated by the nonlexical procedure before lexical retrieval has completed.

These inter-related assumptions about the serial nature of the assembly process and the criterion for initiating articulation have been the focus of recent controversy. Evidence supporting a serial assembly process has been argued to offer a definitive basis for distinguishing between dual route and single-route accounts of word naming (Coltheart & Rastle, 1994; Rastle & Coltheart, 1998, 1999). However, the clarity of this distinction hinges on accepting the DRC assumption that articulation is initiated only when the complete pronunciation has been activated. Recent data demonstrating a variety of systematic effects of the nature of the initial phoneme or phoneme cluster on word naming latency have been argued to demonstrate that subjects do initiate articulation on the basis of initial segments when conditions allow it (Cortese, 1998; Kawamoto, Kello, Higareda & Qu, 1999).

Thus, although the two standard naming tasks differ from the remainder of the task battery in both the fast and slow tails of the RT distribution, neither of these differences allow a definitive conclusion as to whether word frequency effects are underestimated by naming tasks or overestimated by tasks requiring lexical discrimination. But rather than

undermining the utility of the RT analysis approach, the present data highlight the importance of the additional constraints that such data provide. Mean RT data obscure information about the basis of performance differences that are essential to resolving the increasingly fine-grained distinctions between current models of word identification.

Strategic modulation of lexical and nonlexical procedures. Dual route frameworks assume that participants can increase their degree of reliance on nonlexical assembly when they are required to name nonwords as well as words. The present comparison of WNAM and W/NWNAM revealed virtually no differences between the RT distributions for words. This does not necessarily provide evidence against the independence of lexical and nonlexical procedures because the use of predominantly regular words may have provided little incentive for participants to modify their “default” weighting of routes as a function of the presence or absence of nonwords. Including nonwords does seem to induce strategic changes in the processing of “shallow” alphabetic orthographies such as Italian (Tabossi & Laghdi, 1992) and Farsi (Baluch & Besner, 1991), but the evidence in English is very mixed (Monsell et al., 1992 but see Coltheart & Rastle, 1994; Keefe & Neely, 1990; Lupker, Brown & Colombo, 1997).

Comparisons of standard naming tasks with LEXNAM also provide evidence about the independence of lexical and nonlexical procedures. All tasks require identical word responses but the naming tasks yielded RT distributions that differed substantially from the LEXNAM task. Dual route frameworks assume that these differences arise because the LEXNAM task requires participants to disable or ignore outputs of the nonlexical procedure and rely solely on the output of the lexical route. This account predicts greater skew for LF words in LEXNAM than naming tasks and so is indistinguishable from Balota and Spieler’s (1999) claim that a process specific to lexical discrimination tasks yields increased skew in the RT distribution for LF words.

Although this general difference between the LEXNAM and standard naming tasks is consistent with the dual route framework, it is less clear that the model can explain why the RT distribution for LEXNAM yields is virtually identical in shape to that for the LDT but shifted in location by about 120 ms. Differences between the speed of the fastest LDT and LEXNAM responses might be attributed to the time required to retrieve or

assemble the word's pronunciation after lexical retrieval has completed. However, because LF words take longer to retrieve by the lexical procedure, nonlexically assembled pronunciations should be available for many LF words by the time access completes and there should be no need for an additional process following retrieval (Paap et al., 1987). The difference between response times in the LEXNAM and LDT should therefore reduce for longer RTs and be reflected in skew differences paralleling those between LDT and naming tasks. No such differences were observed.

Lexically contingent naming is a relatively recent addition to psychologists' battery of word identification tasks. Forster and Davis (1991) introduced it to evaluate nonlexical influences on the masked priming paradigm and it has recently been used to evaluate the locus of polysemy (Hino & Lupker, 1996) and regularity effects on English and Japanese (Hino & Lupker, 1998). Hino and Lupker (1996, 1998) found that the WFEs on contingent naming were approximately equal to the sum of the LDT and naming WFEs. They therefore suggested that contingent naming involves the processes required for both LDT and word naming and that "individuals essentially carry out these tasks in a sequential order. That is, their decision-making processes are virtually complete before they initiate their pronunciation-related processes" (Hino & Lupker, 1996, p. 1339). Hino and Lupker further assume that WFEs in the LDT and naming tasks reflect different task-specific processes: the LDT effect is due to a decision process similar to that proposed by Balota and Chumbley (1984) while that in the naming task arises from a frequency-sensitive phonological coding process.

The present contingent naming results are not entirely consistent with Hino and Lupker's (1996, 1998). The WFE on mean RT in LEXNAM was slightly larger than that for the LDT but not significantly so, and it certainly did not approximate the sum of the naming and LDT effects. The slower overall RT for LEXNAM is consistent with the view that it requires some additional process beyond that necessary for the LDT. Hino and Lupker argue that this process is similar to that required for standard naming and is frequency-sensitive. But in the present results the latency difference between LEXNAM and the LDT is constant across the RT distribution and equivalent for HF and LF words. The additional process required for LEXNAM does not, therefore, appear to be

influenced by frequency. It might be identified with the time required to retrieve and/or execute an articulatory program following lexical retrieval. This account is similar to Hino and Lupker's because it assumes that contingent naming involves sequential lexical and phonological retrieval, however the present data imply that the contingent naming WFE is entirely due to a lexical process that is essentially the same as that involved in the LDT followed by a frequency-insensitive process that retrieves the pronunciation of the item.

Clearly further data using the contingent naming data are required to resolve the apparent inconsistencies between our results and those of Hino and Lupker (1996, 1998) but there is no necessary conflict between the two data sets. The enhanced WFE in contingent naming tasks reported by Hino and Lupker (1996, 1998) were most marked for polysemous words and for Kanji characters and irregular English words. By contrast, the stimuli in the present experiments were unambiguous and primarily regular and are therefore predicted by Hino and Lupker's account to show a reduced task by frequency interaction. Again, RT distribution analysis will contribute to resolving the basis of the different findings. Information about the differential slowing of fast and slow responses allows a fine-grained evaluation of how the retrieval and pronunciation demands of the contingent naming task are coordinated and how and why they differ as a function of item type.

Single-route models of word naming. The discussion has focused on dual route frameworks of naming performance because they generate more specific and falsifiable hypotheses than single-route accounts. However, the latter can explain much of the evidence taken to demonstrate strategic flexibility. Parallel activation frameworks such as the IA and PDP models do not assume an independent nonlexical procedure, but these models still assume that pronunciations can be assembled from nonlexical constituents to account for nonword naming and often assume that nonlexical and lexical activation have different time-courses that moderate their impact as a function of task context. Monsell et al. (1992) argued that there are two ways of accommodating the functional dissociations claimed to demonstrate strategic variability within PDP frameworks. First, such models could assume "strategy nodes", like the "task nodes" incorporated in Cohen, Dunbar and

McClelland's (1990) simulation of the Stroop task, which modulate the relative weighting of lexical and nonlexical correspondences. Alternatively, they might invoke contributions from the semantic network that act to "clean up" the patterns of activation in orthographic and/or phonological networks. Semantic clean-up has been argued to provide a possible account of the different patterns of performance associated with the various subtypes of acquired and developmental dyslexia (Plaut et al., 1996) and may allow a parallel account of task-based strategy effects (Monsell et al., 1989).

Taking a different tack, Hino and Lupker (1998) claimed that their LDT and contingent naming data are compatible with PDP models precisely because these models do not assume an explicit distinction between lexical and nonlexical representations. Because these models assume that the same network of orthographic units "drive both the original generation of phonological codes in standard naming and the regeneration of phonological codes in ...[contingent] naming" (p. 1448), Hino and Lupker argue that they predict that contingent naming should yield the same pattern of effects as the naming task, but at an inflated magnitude, as they observed for manipulations of both polysemy (Hino & Lupker, 1996) and Japanese script type (Hino & Lupker, 1998). The LDT yields different patterns of effects because it does not require phonological activation (Hino & Lupker, 1996, 1998).

The present results are compatible with this elaboration of the PDP model in the sense that the overall pattern of stimulus effects was consistent across all four naming and lexical discrimination tasks. All showed main effects of frequency and interactions between frequency and animacy due to a larger animacy effect to LF words particularly among slower responses. The magnitude of the WFE, but not the animacy interaction, was enhanced in the lexical discrimination tasks. The qualitative similarity of the distributions for these four tasks seems to implicate a common process, which is compatible with the PDP assumption that they all reflect activation of a common lexical network. However, it is not clear how these models would explain why the animacy effect does not increase in lexical discrimination tasks in parallel with the effects of frequency.

In fact, it is not clear why animacy effects should be observed at all in these tasks. Animacy is irrelevant to either pronunciation or lexical classification, so there is no obvious reason why responses should be faster to animate than inanimate words. The fact that the delayed naming data showed the opposite effect of animacy demonstrates that the animacy effect on pronunciation responses is not due to a confounding with production-related variables which should, if anything, have reduced the advantage for animate over inanimate words in the WNAM, W/NWNAM and LEXNAM tasks.

The effects of the animacy manipulation on the non-semantic judgment tasks may reflect confounding with another variable that influenced lexical retrieval speed. That is, despite the matching of normative frequency and the restriction to concrete nouns, the animate words might be more familiar, or more concrete or imageable than the inanimate words. It is not possible to entirely discount a confounding with familiarity, but if this were the basis of the animacy effects they should have shown a pattern of cross-task differences similar to those of normative frequency – that is, the effects should have been larger in lexical discrimination than naming tasks. Animacy effects were slightly larger in the two lexical discrimination tasks than in WNAM or W/NWNAM and significantly so on some measures. However, the task difference was much smaller than that due to frequency.

Alternatively, the animacy effects may reflect the influence of semantic activation on lexical processing. Whether by virtue of their animacy per se or because of a correlation with semantic variables such as concreteness, top-down feedback might enhance activation of animate words. Such an interpretation is consistent with the fact that animacy effects were more marked for slower responses to LF words because semantic activation would take some time to build up. The SEMCAT task was included to provide insight into semantic activation and judgment processes.

What processes underlie semantic categorization performance?

According to Balota and Spieler's (1999) claim that τ frequency effects reflect a decision strategy specific to lexical classification tasks, semantic categorization tasks should yield RT distributions that are less skewed than those for the LDT, particularly for LF words. By contrast, if the τ frequency effect in the LDT reflects word identification

processes that can sometimes be by-passed for HF words, as implied in models like MROM, then the SEMCAT task should show skew effects for LF words that are equivalent to those observed in the LDT. The SEMCAT data do not correspond completely to either of these extreme alternatives. There are three major differences between the SEMCAT data and that for the other tasks. First, the overall RT distribution for SEMCAT was much more positively skewed than that for any other task: τ was higher for every condition of SEMCAT than for the corresponding condition of the other tasks, including the LF conditions of LDT and LEXNAM. Second, SEMCAT was the only task in which the animacy effect was greater than the effect of frequency. The SEMCAT animacy main effect was reflected in both μ and σ , but not τ while the considerably smaller animacy main effect observed in the other tasks was primarily reflected in τ . Third, whereas all other tasks showed a larger animacy effect for LF words, in SEMCAT the animacy effect was larger for HF words and of similar magnitude across the complete RT distribution. In SEMCAT, there is a clear separation between animate and inanimate words, of both HF and LF, throughout the RT distribution while the RT distributions for animate and inanimate words in all other tasks only diverge for LF words and only as RT increases (see Figure 1).

Examination of the SEMCAT RT and accuracy data suggested that performance was influenced by a bias to make fast “yes” responses to animate words. The accuracy data also suggested a bias to respond “no” to LF words, but it did not lead to the fast errors associated with animacy judgments. Thus, evidence that the item is familiar appears to trigger a fast “animate” response, which leads to a high error rate for HF inanimate words and may yield fast correct “guesses” to LF animate words before semantic information has become available. By contrast, when items are not highly familiar, there is a bias to response “inanimate” which appears to operate after the results of deeper lexical processing – perhaps when this processing does not clearly specify the correct response.

Although supported by a variety of details of the speed and accuracy data, the precise basis of the familiarity bias is unclear. The bias did not appear to depend on carry-over effects from prior experience at lexical classification. The familiarity bias may arise because participants in some sense associate frequency with animacy and tend to classify

frequent words as animate and infrequent words as inanimate. Alternatively, it might occur because familiarity information becomes available earlier in processing than the semantic information that is required to judge animacy so that, in speed-pressured conditions, participants have difficulty inhibiting a response based solely on item familiarity even when it is irrelevant to the task at hand.

Ironically then, although the LDT results provided no direct evidence of a familiarity-based process, the results of the SEMCAT task do appear to implicate a familiarity-based decision mechanism although it is different from that Balota and Chumbley (1984) hypothesized to underlie the LDT. One major difference, of course, is that the animacy judgments required for SEMCAT are not correlated with word frequency so the bias causes errors for inanimate HF words and LF animate words. Secondly, the bias does not appear to yield fast “inanimate” responses to LF words. There is therefore no evidence for an initial stage in which both highly familiar and highly unfamiliar items yield fast responses proposed by Balota and Chumbley’s (1984) account of the LDT.

The further question that needs to be addressed about the semantic categorization task concerns exactly how participants make correct animacy judgments. The results imply that the process underlying such judgments yields faster “animate” than “inanimate” responses. The advantage for “animate” responses might be exaggerated by the familiarity bias, but familiarity is unlikely to provide a complete explanation of the large animacy effect because the average animacy effect on SEMCAT mean RT (92 ms) was larger than the frequency main effect (70 ms), such that classification of animate LF words were faster than for words that were HF but inanimate. Monsell et al. (1989) observed a similar speed advantage for “person” over “thing” judgments. The speed advantage might be taken as evidence that animacy is explicitly coded as a semantic feature, while judgments that an item is inanimate require some more inferential process. For example, Smith, Shoben and Rips (1974) proposed that items that cannot be classified on the basis of global similarity to the relevant category require a more detailed and “deliberative” evaluation of defining semantic features. However, such a view seems to predict a greater proportion of slow responses to inanimate than animate words which

should be reflected in differential skew. In fact, the differences between animate and inanimate words were relatively constant across the RT distribution and so reflected primarily in μ . This implies that the animacy effect reflects an additional process that adds a constant increment to “inanimate” responses. This seems most parsimoniously identified with the speed advantage for positive over negative decisions (e.g., Sternberg, 1969) and, like the familiarity bias, implies that participants treated the task as requiring a “yes/no” judgment as to whether the item referred to an animate object. By this view, correct semantic categorization judgments that are not based on familiarity rely on lexical and semantic retrieval processes that are similar for animate and inanimate words, followed by a binary decision and/or response related process that is faster for “yes” than “no” responses.

Thus, semantic categorization does not provide as pure a measure of lexical or semantic retrieval as has been assumed (e.g., Balota & Chumbley, 1984; Forster & Shen, 1996). Participants are capable of making very accurate judgements of animacy demonstrating that semantic information is effectively retrieved for most items. However, they seem to be unable to avoid the influence of familiarity-based information even though it is not a reliable guide to the required response. In this sense, the semantic categorization task appears to be just as susceptible to task-specific decision strategies as LDT and naming tasks and the WFE in this task provides no more valid an index of the role of word frequency in lexical access than any other word identification task. This means that the demonstration of equivalent WFEs on mean RT in SEMCAT and the LDT cannot be taken as evidence that they rely on a common access procedure as has been assumed within the “conventional” view of lexical access (Monsell et al., 1989). But equally, the absence of a WFE in semantic categorization data as reported by Balota and Chumbley (1984) does not prove that frequency does not influence lexical retrieval. Each task needs to be analyzed to determine the relative contributions of general lexical retrieval and specific strategies that allow participants to generate the required response. RT distribution analyses provide a sensitive methodology for conducting this analysis.

What is the relationship between familiarity and word identification?

The issue most central to interpreting both the present results and the two earlier investigations of RT distributions in word identification tasks concerns the processes that are indexed by τ . Balota and Spieler (1999) attributed the τ component of the WFE to a decision process specific to the LDT. Although the present results replicate this finding, they show that high τ is also characteristic of the SEMCAT task, which yielded considerably larger τ estimates than all other tasks in every stimulus condition. The τ effects in the other tasks were larger for LF words, so the differential skew in SEMCAT is most marked for HF words.

Thus, the smaller τ frequency effect in SEMCAT compared to LDT and LEXNAM occurs not because τ is larger for LF words in lexical discrimination tasks – as would be expected if τ reflects an additional decision process – but rather because HF words were associated with a larger τ in SEMCAT than in the other two tasks. Framed in this way, τ might well be identified with lexical retrieval processes that can sometimes be by-passed for HF words in the LDT and for LF regular words in naming tasks. LDT responses to HF words can sometimes be made before lexical retrieval has completed because the item is high in familiarity or elicits high global activation. This might account for why HF words show less skew in LDT than SEMCAT. Naming responses are sometimes generated nonlexically, particularly for LF words, so naming tasks show reduced skew for LF words compared to tasks that preclude reliance on nonlexical assembly.

The present results do not provide a definitive basis for distinguishing between Balota and Spieler's (1999) claim that τ reflects a strategy specific to lexical discrimination tasks and the alternative possibility that it indexes a lexical retrieval process common to a variety of word identification tasks. The enhanced τ component of the WFE combined with the fact that WFE effects in SEMCAT manifested in μ rather than τ , is consistent with Balota and Spieler's (1999) claim that τ reflects a process specific to lexical discrimination. However, the fact that τ was greater for LF than HF words in the naming and LEXNAM tasks as well as the LDT, and that τ was high for all conditions of SEMCAT, suggests that it reflects a general rather than task-specific process.

It is important to acknowledge that Balota and Spieler's (1999) interpretation of their data did not rely solely on comparisons of the WFE in LDT and naming. They also investigated the effects of stimulus repetition and its interaction with frequency because the combined effects of frequency and repetition provide strong constraints that can be used to distinguish between different models of lexical retrieval. Balota and Spieler's (1999) difficulty in formally simulating their data with different two-stage models arose primarily from the joint effects of these variables. However, it is clear that repetition effects do not reflect a unitary mechanism (e.g., Forster & Davis, 1984; Monsell, 1985). Thus, as Balota and Spieler's final hybrid model exemplifies, complex relationships between frequency and repetition effects can be accommodated by assuming that repetition has independent effects on different word identification processes. Such an assumption makes it difficult to draw definitive conclusions about the processes implicated by a particular set of behavioral outcomes. Balota and Spieler's approach demonstrates the benefits of RT distributional analysis in providing additional constraints that can contribute to testing distinctions between different combinations of assumptions. Further investigations of the effects of repetition on RT distributions obtained in a wider variety of tasks are necessary to evaluate the validity of their hybrid model.

The more general conceptual issue underlying Balota and Spieler's (1999) theoretical approach concerns whether the mechanism responsible for familiarity effects is independent of the lexical retrieval process, as implied in Balota and Chumbley's (1984) original two-stage model. Monsell et al. (1989) pointed out that a familiarity estimate of the form described by Balota and Chumbley could be derived from on-going lexical retrieval rather than being computed extra-lexically. This is the assumption incorporated in the MROM and DRC models, but these models still assume that this information is only recruited for the LDT in which it provides an alternative basis for a classification response which is artifactual to "true" lexical retrieval (Grainger & Jacobs, 1996).

It is not, however, necessary to assume that familiarity and lexical identification exert separate influences on performance. Reichle, Pollatsek, Fisher and Rayner (1998) have recently proposed a model of word identification in which familiarity estimates

contribute to the normal word identification process. This model was developed to account for the pattern of eye movements during reading and, in particular, for evidence suggesting that different mechanisms underlie the programming of saccadic eye movements to a new word in the text and the shift of covert attention to that word. Reichle et al. (1998) argue that such evidence indicates that lexical retrieval is not “the engine that drives eye movements” (Reichle et al., 1998, p. 129) as assumed by earlier models of eye movement control (e.g., Morrison, 1984). Their new model therefore “decouples covert attention shifts from eye movement programming” (p. 133) by distinguishing between the early computation of a word’s familiarity and the ultimate retrieval of a single lexical representation. Because familiarity provides an index of “the proficiency and probability of successful resolution” of the retrieval process it serves as a “signal for the initiation of an eye movement program [while]... completion of lexical access [is] the signal for a shift of covert attention” (p. 133). Critically, in the present context, Reichle et al. (1998) do not attribute familiarity and lexical retrieval effects to different mechanisms. Both are computed by a single “word recognition module”, and both are assumed to be affected by many of the same factors. They point to the parallel between the familiarity/ lexical retrieval distinction and the distinction between “matching on global similarity” and “retrieval through reintegration, or pattern completion” that has been incorporated into a number of models of memory including MINERVA (Hinzman, 1988), CHARM (Eich, 1985), SAM (Gillund & Shiffrin, 1984) and the matrix model (Humphreys, Bain & Pike, 1989). In such models, a memory probe can either be compared with either the collective contents of memory to produce a scalar value that represents the item’s global familiarity or similarity to other memory traces or used to retrieve a vector representing the content of the memory representation of the item. The retrieval process is more difficult and computationally demanding than the matching process because it requires retrieval of unique item information, but both processes are derived from a single memory system. Similarly, “the familiarity check and lexical access could be the product of the same word recognition module” (Reichle et al., 1998, p. 133).

The relevance of Reichle et al.'s model to the present discussion is that it provides a functional explanation of the role of familiarity in normal word identification. If the only purpose in computing a measure of lexical familiarity early in lexical processing were to allow participants to make quick LDT responses, then it would be reasonable to view this computation as essentially artifactual with respect to "normal word identification". However, according to Reichle et al.'s (1998) rationale, deriving an early estimate of the probability that the word will be successfully retrieved is crucial to the efficient coordination of eye movements during normal reading. This estimate might also be co-opted to facilitate performance in laboratory word identification tasks like the LDT. The apparent contribution of familiarity to performance in the semantic categorization task, where it is irrelevant to the required judgment, is consistent with the view that familiarity is computed as part of the normal word identification process and influences performance even when it leads to errors.

Recognizing that familiarity-related processes may play an important role in normal word identification might also explain the qualitative similarity of the RT distributions obtained in the lexical discrimination and naming tasks. Despite differences in overall RT and the magnitude of the τ frequency effect in the two task clusters, all four naming and lexical discrimination tasks showed increasing effects of both word frequency and animacy across the RT distribution. As discussed in relation to PDP models of word naming, parallel effects of these stimulus variables on the RT distributions for discrimination and naming task is consistent with the involvement of a common process that is required for all tasks. However, the different response requirements associated with word naming and lexical classification appears to either "compress" the RT distribution associated with naming or "stretch" the distribution for the lexical discrimination tasks. Within Reichle et al.'s (1998) model, both types of task might be assumed to rely on a similar evidence accumulation process that first gives rise to a measure of overall familiarity and gradually provides the more detailed information associated with lexical retrieval, which is reflected in the increasing animacy effects with RT observed in both tasks. Differences between the tasks might then be attributed to the nature and efficiency of the processes that make use of this information to produce the

required response. People's practice with the task of reading words aloud may allow them to generate the appropriate response quickly and efficiently from the outputs of the word identification process, while the unfamiliar requirement of making button-press classifications about item lexicality requires greater refinement of the evidence to generate a response, particularly for LF words. The contribution of these differential response-related processes is reflected in the relative compression and stretching of the RT distributions obtained in the naming and lexical discrimination tasks.

Conclusions

The results replicated virtually all of Balota and Spieler's (1999) findings about the differential effects of word frequency on RT distributions for naming and LDT tasks. However, the data provided by a broader array of tasks and more refined evaluations of the distributional data do not unequivocally support their conclusion that the large WFE obtained in the LDT reflects a task-specific decision process. The present pattern of data is equally compatible with the view that all of the tasks included in the current battery rely on a common, frequency-sensitive, word identification process, but that this process can sometimes be by-passed or reduced for HF items in the LDT because of the influence of familiarity-based processes; and for LF words in naming tasks because of the contribution of nonlexical assembly mechanisms. Such an analysis is potentially compatible with a variety of frameworks including dual route, IA and PDP models although all have difficulty dealing with some details of the data.

The difficulty in conclusively distinguishing between the predictions of conceptually distinct models reflects the fact that the precision of the data that we have presented exceeds, in many respects, the precision of the models that they were collected to test. It has been possible to draw out specific predictions about RT distributions from a variety of models, but it is often difficult to determine exactly how these predictions would be manifested within the complex hybrid models that are dominating the word recognition literature. Computational models are proliferating as quickly as the power of computers that allow them to be developed, so it is increasingly important to provide tighter empirical constraints that can be used to discriminate between overlapping models and determine the relative validity of different conceptual approaches. The present data

demonstrate the potency of analyses of RT distributions as a source of such empirical constraint and as a stimulus to development of more comprehensive and psychologically valid models.

References

- Andrews, S. (1982) Phonological recoding: Is the regularity effect consistent? Memory & Cognition, 10, 565-575.
- Andrews, S. (1997) The role of orthographic similarity in lexical retrieval: resolving neighborhood conflicts. Psychonomic Bulletin and Review, 4, 439-461.
- Atkinson, R.C. & Juola, J.F. (1974) Factors influencing speed and accuracy of word recognition. In S. Kornblum (Ed.), Attention and Performance IV (pp. 583-612). New York: Academic Press.
- Balota, D. A. (1990) The role of meaning in word recognition. In D. A. Balota, Flores d'Arcais, G. B. & Rayner, K. (Eds.), Comprehension processes in reading. Hillsdale, NJ: Erlbaum.
- Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. Journal of Experimental Psychology: Human Perception and Performance, 10, 340-357.
- Balota, D. A., & Chumbley, J. I. (1985). The locus of word-frequency effects in the pronunciation task: Lexical access and/or production? Journal of Memory and Language, 24, 89-106.
- Balota, D. & Abrams, R. (1995) Mental chronometry: Beyond onset latencies in the lexical decision task. Journal of Experimental Psychology: Learning, Memory and Cognition, 21, 1289-1302.
- Balota, D.A. & Shields, L. (1988) Localizing word frequency effects in pronunciation. Paper presented at the annual meeting of the Psychonomic Society, November, Chicago.
- Balota, D.A. & Spieler, D.H. (1997) Bringing computational models of word naming down to the item level. Psychological Science, 8, 411-416.
- Balota, D. A. & Spieler, D. H. (in press) Word frequency, repetition and lexicality effects in word recognition tasks: Beyond measures of central tendency. Journal of Experimental Psychology: General.

Baluch, B. & Besner, D. (1991) Visual word recognition: Evidence from the strategic control of lexical and nonlexical routines in oral reading. Journal of Experimental Psychology: Learning, Memory and Cognition, 17, 644-652.

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

Becker, C. A. & Killion, T. H. (1977) Interaction of visual and cognitive effects in word recognition. . Journal of Experimental Psychology: Human Perception and Performance, 3, 389-401.

Besner, D. & Johnston, J.C. (1987) Reading and the mental lexicon: On the uptake of visual information. In W. Marslen-Wilson (Ed.), Lexical representations and processes. Cambridge, Mass.: MIT Press (pp. 291-316).

Borowsky, R. & Besner, D. (1993) Visual word recognition: A multistage activation model. Journal of Experimental Psychology: Learning, Memory and Cognition, 19, 815-840.

Burbeck and Luce (1982). Evidence from auditory simple reaction times for both change and level detectors. Perception and Psychophysics, 32, 117-133.

Cohen, J.D., Dunbar, K. & McClelland, J.L. (1990) On the control of automatic processes: A parallel distributed processing account of the Stroop effect. Psychological Review, 97, 332-361.

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.

Coltheart, M., & Rastle, K. (1994). Serial processing in reading aloud: Evidence for dual route models of reading. Journal of Experimental Psychology: Human Perception and Performance, 20, 1197-1211.

Coltheart, M., Rastle, K., Perry, C., Langdon, R. & Ziegler, J. (in press) DRC: A dual route cascaded model of visual word recognition and reading aloud. Psychological Review.

Connine, C., Mullenix, J., Shernoff, E. & Yelens, J. (1990) Word familiarity and frequency in visual and auditory word recognition. Journal of Experimental Psychology: Learning, Memory and Cognition, 16, 1084-1096.

Cortese, M.J. (1998) Revisiting serial position effects in reading. Journal of Memory and Language, 39, 652-665.

Cox, D. R. & Hinckley, D. V. (1974) Theoretical statistics. London: Chapman and Hall.

Efron, B. & Tibshirani, R. J. (1993) An introduction to the bootstrap. NY: Chapman and Hall.

Eich, J. M. (1985) A composite holographic associative recall model. Psychological Review, 89, 627-661.

Forster, K. I. (1976) Accessing the mental lexicon. In R. Wales & E. Walker (eds.), New approaches to language mechanisms. Amsterdam: North Holland.

Forster, K. I. (1985) Lexical acquisition and the modular lexicon. Language and Cognitive Processes, 1, 87-108.

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

Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. Journal of Experimental Psychology: Learning, Memory and Cognition, 10, 680-698.

Forster, K.I. & Davis, C. (1991) The density constraint on form-priming in the naming task: Interference effects from a masked prime. Journal of Memory and Language, 30, 1-25.

Forster, K. I. & Shen, D. (1996) Neighborhood frequency and density effects in visual word recognition. Journal of Experimental Psychology: Learning, Memory and Cognition, 22, 606-713.

Frederiksen, J. R. & Kroll, J. F. (1976) Spelling and sound: Approaches to the internal lexicon. Journal of Experimental Psychology: Human Perception and Performance, 2, 361-379.

Gillund, G. & Shiffrin, R.M. (1984) A retrieval model for both recognition and recall. Psychological Review, 91, 1-67.

Glanzer, M. & Bowles, N. (1976) Analysis of the word-frequency effect in recognition memory. Journal of Experimental Psychology: Human Learning and Memory, 2, 21-31

Goldinger, S.D., Azuma, T., Abramson, M. & Jaim, P. (1997) Open wide and say "Blah!": Attentional dynamics of delayed naming. Journal of Memory and Language, 37, 190-216.

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

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

Heathcote, A. (1996) RTSYS: A computer program for analysing response time data. Behaviour Research Methods, Instruments & Computers, 28, 427-445.

Heathcote, A., Popiel, S.J. & Mewhort, D. J. K. (1991) Analysis of response time distributions: An example using the Stroop task. Psychological Bulletin, 109, 340-347.

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

Hino, Y. & Lupker, S.J. (1998) The effect of word frequency for Japanese Kana and Kanji words in naming and lexical decision: Can the dual-route models save the lexical-selection account? Journal of Experimental Psychology: Human Perception and Performance, 24, 1431-1453.

Hintzman, D. L. (1988) Judgments of frequency and recognition memory in a multiple-trace model. Psychological Review, 95, 528-551.

Hockley, W. E. (1984) Analysis of response time distributions in the study of cognitive tasks. Journal of Experimental Psychology: Learning, Memory and Cognition, 10, 598-615.

Hohle, R. H. (1965) Inferred components of reaction times as functions of foreperiod duration. Journal of Experimental Psychology, 69, 382-386.

Humphreys, M. S., Bain, J. D. & Pike, R. (1989) Different ways to cue a coherent memory system: A theory for episodic, semantic and procedural tasks. Psychological Review, 96, 208-233.

Kawamoto, A.H., Kello, C.T., Higareda, I. & Qu, J.V. (1999) Parallel processing and initial phoneme criterion in naming words: Evidence from frequency effects on onset and rime duration. Journal of Experimental Psychology: Learning, Memory and Cognition, 25, 362-381.

Kawamoto, A. H. & Zemplidige, J. (1992) Pronunciations of homographs. Journal of Memory and Language, 31, 349-374.

Keefe, D.E. & Neely, J. H. (1990) Semantic priming in the pronunciation task: The role of prospective prime generated expectancies. Memory & Cognition, 18, 289-298.

Kucera, H. & Francis, W. N. (1967) Computational analysis of present-day American English. Providence, R.I.: Brown University Press.

Luce, R. D. (1986) Response times: Their role in inferring elementary mental organization. NY: Oxford University Press.

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

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. Psychological Review, 88, 375-407.

McCrae, K., Jared, D. & Seidenberg, M.S. (1990) On the roles of frequency and lexical access in word naming. Journal of Memory and Language, 29, 43-65.

Mewhort, D.J.K., Braun, J. G. & Heathcote, A. (1992) Response time distributions and the Stroop task: A test of the Cohen, Dunbar and McClelland (1990) model. Journal of Experimental Psychology: Human Perception and Performance, 18, 872-882.

Miller, J. O. (1991) Reaction time analysis with outlier exclusion: Bias varies with sample size. Quarterly Journal of Experimental Psychology, Quarterly Journal of Experimental Psychology, 43A, 907-912.

Monsell, S. (1985) Repetition and the lexicon. In A.W. Ellis (Ed.), Progress in the Psychology of Language (Vol. 2) (pp. 147-195), London: Erlbaum.

Monsell, S., Doyle, M. C. & Haggard, P. N. (1989) Effects of frequency on visual word recognition tasks: Where are they? Journal of Experimental Psychology: General, 118, 43-71.

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

Morrison, R. E. (1984) Manipulation of stimulus onset delay in reading: Evidence for parallel programming of saccades. Journal of Experimental Psychology: Human Perception and Performance, Journal of Experimental Psychology: Human Perception and Performance, 10, 667-682.

Morton, J. (1970). A functional model for memory. In D. A. Norman (Ed.), Models of Human Memory Academic Press.

Paap, K. E., McDonald, J. E., Schvaneveldt, R. W. & Noel, R. W. (1987) Frequency and pronounceability in visually presented naming and lexical-decision tasks. In M. Coltheart (Ed.), Attention and performance XII (pp. 221-244). Hillsdale, NJ: Erlbaum.

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

Plourde, C. E. & Besner, D. (1997) On the locus of the word frequency effect in visual word recognition. Canadian Journal of Experimental Psychology, 51, 181-194.

Rajaram, S. & Neely, J. H. (1992) Dissociative masked repetition priming and word frequency effects in lexical decision and episodic recognition tasks. Journal-of-Memory-and-Language, 31, 152-182.

Ratcliff, R. (1979) Group reaction time distributions and an analysis of distribution statistics. Psychological Review, 86, 446-461.

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

Rastle, K. & Coltheart, M. (1998) Whammy and double whammy: Length effects in nonword naming. Psychonomic Bulletin and Review, 5, 277-282.

Rastle, K. & Coltheart, M. (1999) Serial and strategic effects in reading aloud. Journal of Experimental Psychology: Human Perception and Performance, 25, 482-503.

Reichle, E. D., Pollatsek, A., Fisher, D. L. & Rayner, K. (1998) Toward a model of eye movement control in reading. Psychological Review, 105, 125-157.

Savage, G.R., Bradley, D.C. & Forster, K.I. (1990) Word frequency and the pronunciation task: The contribution of articulatory fluency. Language and cognitive processes, 5, 203-236.

Scarborough, D.L., Cortese, C. & Scarborough, H.C. (1977) Frequency and repetition effects in lexical memory. Journal of Experimental Psychology: Human Perception and Performance, 3, 1-17.

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

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

Smith, E., Shoben, E. & Rips, T. (1974) Structure and process in semantic memory: A featural model of semantic decisions. Psychological Review, 81, 214-241.

Spieler, D.H., Balota, D.A. & Faust, M.E. (1996) Stroop performance in normal older adults and individuals with senile dementia of the Alzheimer's type. Journal of Experimental Psychology: Human Perception and Performance, 22, 461-479.

Sternberg, S. (1969) The discovery of processing stages: Extension of Donders' methods. Acta Psychologica, 30, 276-315.

Stone, G. O. & Van Orden, G. C. (1993) Strategic control of processing in word recognition. Journal of Experimental Psychology: Human Perception and Performance, 19, 744-774.

Tabossi, P. & Laghdi, L. (1992) Semantic priming in the pronunciation of words in two writing systems: Italian and English. Memory & Cognition, 20, 303-313.

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

Vincent, S. B. (1912) The function of the vibrissae in the behavior of the white rat. Behavioral Monographs, 1 (No. 5).

Weekes, B. S. (1997) Differential effects of number of letters on word and nonword naming latency. Quarterly-Journal-of-Experimental-Psychology, 50A, 439-456.

Whaley (1978) Word-nonword classification time. Journal of Verbal Learning and Verbal Behavior, 17, 143-154.

Wilkinson, L. (1999). Statistical methods in psychology journals: guidelines and explanations. American Psychologist, 54, 594-604.

Author note

This research was supported by Australian Research Council Grant A79601998 to Sally Andrews and reported at the Psychonomic Society Conference held in November, 1998 in Dallas. The manuscript was prepared while the first author was a sabbatical visitor at the Universite Libre de Bruxelles and the University of Massachusetts, Amherst and she is grateful to Alain Content and Keith Rayner for providing a stimulating environment and access to facilities to achieve this task. Monica Blayney, Colin Davis and Danielle R. Scarratt made valuable contributions to data collection and collation and David Balota, Stephen Lupker and David Plaut made constructive criticisms of earlier versions of the manuscript. We would like to express our appreciation to Don McNicol for starting both of us off on paths that ultimately converged on this collaboration.

Appendix A

Listing of the mean RT for each word in each of the tasks included in Experiment 1.

WORDS	NAM	NWNAM	LEXNAM	LDT	SEMCAT	NAM	NWNAM	LEXNAM	LDT	SEMCAT	
	High Frequency Animate					Low frequency Animate					
artist	502	579	614	480	660	alderman	607	705	771	630	653
bride	525	620	603	506	645	assassin	573	733	784	658	720
buffalo	516	587	601	579	718	attacker	513	652	648	523	734
candidate	589	691	715	666	762	bitch	501	610	609	485	679
carpenter	556	643	695	524	700	bluebird	561	618	610	533	709
chairman	611	660	739	532	647	boar	543	667	748	590	684
consumer	563	664	718	561	747	cheetah	636	748	783	631	677
cow	530	599	644	523	600	chef	601	708	776	535	710
donkey	499	616	553	548	659	cleric	637	824	758	685	831
editor	484	576	647	530	667	dolphin	532	605	631	573	592
human	647	673	675	535	573	foal	561	768	819	591	863
insect	496	632	580	532	707	guru	626	712	842	687	742
king	536	606	640	494	702	juror	697	761	779	689	705
kitten	531	595	680	495	665	kidnapper	572	670	686	548	637
maid	469	559	585	476	677	mermaid	500	576	574	567	683
mayor	502	593	655	472	713	playboy	526	623	685	520	655
officer	479	608	618	560	707	python	578	666	745	597	713
pig	531	578	607	504	595	ruffian	595	683	698	758	958
rabbit	484	566	565	541	642	sentry	635	759	877	688	739
rat	459	522	562	520	626	spouse	653	732	786	573	797
senator	622	761	752	553	730	stallion	640	759	839	589	692
son	570	626	708	492	615	suitor	640	712	819	624	855
thief	560	630	683	520	677	teenager	553	627	665	527	629
tiger	537	596	615	494	600	tortoise	571	692	709	581	680
worm	501	554	658	504	693	umpire	548	692	712	572	748
ant	554	522	710	518	684	admirer	592	575	683	560	712
assistant	612	590	620	527	734	analyst	565	555	751	535	796
colonel	807	693	911	654	759	anteater	705	713	880	768	910
cousin	569	558	631	479	647	auditor	614	559	730	548	840
deer	560	526	633	564	659	boatman	585	625	731	550	720
doctor	507	502	571	501	575	caller	606	558	776	537	774
dragon	539	564	603	447	748	caretaker	639	605	686	540	687
eagle	542	547	598	509	635	cobra	595	601	722	504	680
engineer	530	547	587	517	705	cockatoo	591	648	679	547	676
female	582	572	647	491	587	commuter	639	631	839	703	773
fox	554	534	635	507	607	gazelle	624	557	729	599	848
frog	559	550	686	486	726	grandson	583	578	661	591	715
gentleman	568	593	654	492	666	inmate	653	647	825	663	1048
goose	549	507	665	494	689	migrant	558	528	638	563	762
horse	509	537	662	468	613	nudist	604	564	707	539	715
lawyer	553	516	652	462	653	parson	624	584	868	588	855
passenger	548	556	645	474	692	plumber	597	570	655	536	636
pilot	567	569	643	500	636	poodle	599	576	736	519	670
poet	578	539	655	472	704	rogue	554	561	711	576	905
princess	564	584	663	511	674	shrimp	596	617	812	574	775
rider	521	508	625	458	803	squaw	746	686	930	736	962
teacher	566	576	606	513	600	stag	614	587	775	589	890

traveller	552	546	671	467	648	steward	711	664	814	510	724
turtle	559	546	684	514	565	walrus	604	588	725	658	731
wife	490	539	621	455	571	weasel	527	526	668	560	803
adult	591	544	585	474	563	addict	649	594	677	546	713
bird	548	525	544	479	714	antler	660	660	745	648	815
chief	716	793	698	482	773	burglar	622	600	625	542	641
clerk	633	586	666	504	667	catfish	621	574	673	531	751
companion	681	672	638	528	631	celebrity	706	689	740	539	662
composer	690	607	685	570	717	columnist	712	710	745	596	776
cult	642	596	649	544	1052	doorkeeper	626	642	641	605	713
customer	586	639	596	523	581	drake	606	603	697	638	770
duke	627	591	649	559	772	florist	635	614	608	481	815
elephant	562	584	578	486	694	gunman	617	586	661	634	634
fairy	610	597	648	507	705	hostage	565	565	610	533	662
girl	572	547	579	460	514	janitor	635	619	645	583	675
hunter	575	530	599	494	609	jockey	592	570	639	533	620
knight	562	585	607	477	587	layman	590	580	659	623	605
lion	552	540	601	483	623	mammal	653	611	627	529	803
monkey	527	520	561	464	596	midwife	571	533	584	537	696
negro	617	555	644	612	681	motorist	566	602	601	560	686
owl	565	576	627	482	615	nun	556	516	593	582	595
pupil	591	592	605	526	697	peacock	581	579	594	485	787
servant	641	679	643	473	586	ranger	552	568	608	535	646
soldier	633	585	600	502	539	slug	656	659	671	532	893
uncle	557	551	574	477	611	sniper	639	653	784	670	745
visitor	574	555	585	460	565	terrier	688	675	707	565	692
wolf	517	518	533	482	715	vicar	653	666	688	554	745
worker	499	562	538	511	617	watchdog	568	561	609	541	633
angel	619	565	646	502	570	boa	588	606	852	842	755
aunt	568	563	660	574	567	bodyguard	577	578	666	550	651
baker	524	559	612	510	616	cadet	625	645	762	608	803
bishop	599	578	634	576	721	camper	570	558	652	550	753
captain	602	539	622	473	555	cellist	749	883	799	642	894
child	565	568	623	502	550	codfish	626	617	799	697	681
citizen	600	639	720	515	581	craftsman	653	649	762	530	602
commander	634	653	741	593	616	crusader	604	667	776	566	697
creature	563	607	694	517	791	cyclist	678	670	744	496	630
farmer	574	567	646	480	624	flea	597	572	689	583	671
grandmother	596	596	702	573	582	gnome	614	646	717	610	650
guard	570	557	611	467	587	grub	542	556	731	663	834
hawk	585	571	668	538	699	hawker	567	560	698	574	776
hero	546	522	617	539	569	jaguar	659	651	780	578	772
lady	520	518	545	480	521	lass	525	501	638	613	716
lamb	490	530	637	440	607	lioness	599	619	670	599	696
painter	554	579	608	480	562	machinist	707	674	695	553	774
parent	562	597	648	508	628	mammoth	586	633	715	609	768
pony	564	544	624	565	638	nymph	561	542	668	625	862
priest	555	599	642	496	621	octopus	539	565	595	573	613
principal	557	578	678	581	696	referee	574	624	654	606	746
sheriff	579	629	733	543	654	refugee	617	591	686	613	709
singer	564	597	731	556	541	seahorse	653	674	790	560	674
snake	594	613	706	467	646	waiter	515	502	590	503	661
student	641	657	721	461	548	wizard	540	520	609	509	617
actor	529	515	574	517	588	aviator	690	589	746	664	729
animal	528	536	607	506	618	beatnik	800	713	818	797	945

author	532	557	597	561	582	bison	709	643	750	687	732	
baby	534	533	535	488	547	culprit	624	650	729	627	764	
beast	640	558	582	527	702	dingo	562	616	631	673	600	
bull	546	571	628	528	591	doe	572	581	769	602	660	
chicken	587	590	622	493	586	drummer	609	603	683	556	692	
crow	620	618	690	561	629	eel	618	565	738	553	700	
daughter	539	579	599	527	527	elk	539	547	690	630	892	
dog	530	515	603	510	581	gangster	590	612	672	577	620	
duck	558	569	591	542	588	golfer	535	540	658	583	571	
emperor	635	633	738	583	630	grocer	607	626	660	571	674	
guest	560	561	612	548	622	gymnast	626	638	709	576	645	
mortal	577	555	652	658	760	heroine	621	607	707	628	739	
prisoner	609	604	693	474	587	hyena	734	641	742	583	795	
sailor	606	637	674	553	532	intruder	554	554	592	590	671	
salmon	699	681	877	554	811	lecturer	553	558	649	617	636	
scientist	615	634	724	546	602	locust	658	611	783	585	769	
sergeant	758	632	766	580	666	logger	607	577	660	635	733	
sister	561	588	667	543	551	newt	675	565	956	728	785	
speaker	661	635	688	560	674	orphan	534	550	661	503	578	
stranger	674	700	735	564	589	pheasant	773	816	935	570	743	
visitor	545	547	573	511	608	prostitute	718	649	690	656	659	
widow	575	547	605	591	685	rosella	617	605	820	647	741	
woman	528	536	595	552	550	shopper	571	553	661	487	633	
		High frequency Inanimate						Low frequency Inanimate				
axe	498	630	623	556	755	amplifier	558	634	724	602	774	
button	492	601	615	516	731	canteen	586	655	660	558	757	
combat	548	636	652	576	766	cosmetic	557	716	692	530	740	
commerce	620	705	802	600	760	dogma	553	695	652	829	776	
compass	551	667	767	540	745	facade	695	710	760	672	755	
cord	541	592	685	542	736	gash	539	591	717	670	685	
curse	571	655	667	532	744	gazette	547	635	703	573	778	
evil	487	563	566	512	775	greed	511	617	656	616	807	
fire	535	592	627	504	650	irony	548	602	671	531	760	
furniture	573	658	655	558	687	kiosk	629	702	724	637	759	
gate	500	578	594	497	687	lattice	542	616	744	628	776	
honey	509	596	622	543	884	leash	511	535	707	649	807	
interval	525	624	637	575	836	limbo	494	592	667	612	878	
island	508	582	627	505	772	loophole	591	758	741	689	764	
jet	505	572	600	486	759	lotion	576	706	753	596	767	
job	489	607	626	478	756	mascara	563	681	679	649	783	
juice	549	652	642	504	817	mead	500	584	664	696	847	
kitchen	537	583	712	494	736	pamphlet	602	673	745	594	751	
label	513	539	581	569	735	passport	551	665	720	540	702	
location	588	656	724	559	687	saxophone	623	674	771	585	808	
mirror	495	567	564	504	707	scoop	670	722	836	525	777	
mystery	480	585	580	498	693	tapestry	563	646	695	562	726	
oxygen	505	574	635	572	737	tarpaulin	740	829	872	753	803	
rocket	470	552	612	496	881	trilogy	668	792	792	600	744	
vegetable	504	594	639	573	724	turban	570	655	749	692	788	
brass	576	537	674	557	821	brawl	555	598	700	587	768	
calm	597	539	691	572	815	calico	701	715	821	683	926	
crime	591	566	672	487	763	chutney	659	667	828	640	964	
ditch	561	532	729	556	798	curry	648	592	825	566	731	
drug	524	505	616	491	826	gutter	553	528	739	548	771	
example	545	581	616	483	790	kennel	632	660	757	527	919	

fan	550	577	673	478	704	kerchief	850	755	963	738	1047
fog	594	574	764	583	857	lounge	559	540	635	551	700
laboratory	638	582	767	549	867	manor	585	579	810	578	1031
laughter	573	548	664	536	862	microwave	552	552	606	530	795
leather	542	566	600	530	776	minuet	679	661	912	703	1085
linen	597	575	752	550	753	obelisk	720	743	815	782	897
liquid	515	549	601	569	769	outburst	595	596	736	561	899
list	493	499	613	481	712	pharmacy	671	661	711	588	882
machine	493	573	581	497	766	poppy	586	565	688	527	945
medicine	550	542	656	574	799	saucepan	580	589	749	507	881
notice	576	494	599	559	746	seafood	632	594	680	509	816
parade	599	571	666	493	793	serenade	687	640	877	669	958
paragraph	643	644	749	563	797	shoelace	658	715	790	654	816
pepper	547	546	631	521	751	shrapnel	673	708	903	724	789
poem	568	573	665	535	796	soccer	643	592	725	498	906
prayer	568	575	708	477	914	spout	686	669	917	568	960
rubber	528	503	596	492	822	stilts	731	702	940	793	850
symphony	651	635	759	559	787	stretcher	648	679	825	582	968
universe	553	546	625	581	861	vaccine	644	682	722	552	940
ambition	632	620	632	604	723	chess	601	658	644	522	699
ballet	645	627	734	524	922	crate	619	612	722	572	688
bath	546	570	605	440	708	drizzle	591	619	673	584	794
canoe	639	666	707	569	761	gadget	640	615	603	602	740
consonant	869	750	975	652	804	gymnasium	721	687	713	605	783
core	606	612	676	547	719	ion	718	691	975	704	779
cottage	608	590	605	488	664	lacquer	720	703	784	627	755
dairy	634	593	656	481	892	ledger	575	566	688	566	811
deck	562	546	581	514	712	limousine	618	643	626	610	750
dirt	560	585	632	511	635	linoleum	871	753	786	657	773
door	548	536	559	431	695	mailbox	545	529	565	537	710
edge	585	528	550	493	765	manual	558	621	632	507	817
expense	643	611	740	591	709	nudge	544	538	614	619	821
flame	627	604	649	539	782	oatmeal	621	640	724	573	744
height	559	620	646	521	801	option	646	543	592	525	777
hut	590	585	685	485	727	paddock	710	637	697	521	854
laugh	541	495	564	488	821	parasol	738	710	817	699	843
library	541	561	623	497	699	pendant	695	689	745	580	838
noise	528	510	541	464	769	plagiarism	794	722	834	715	852
ocean	557	552	575	525	708	pothole	776	682	732	715	752
pillow	580	585	600	528	659	proton	687	671	817	776	841
powder	614	610	625	539	709	resin	675	598	860	726	802
stove	638	652	676	511	784	retina	606	639	667	559	722
syllable	750	694	737	586	704	shuffle	668	602	678	572	773
wedding	531	569	581	483	712	smallpox	767	623	754	561	850
arch	598	597	659	539	710	bagpipe	618	626	740	634	813
battle	522	564	635	537	652	beacon	586	568	686	639	840
bottle	549	545	561	463	660	dinghy	687	690	775	755	753
cave	589	573	692	582	610	easel	603	543	890	544	902
church	578	597	673	509	625	itch	522	532	636	646	737
cloud	571	560	686	503	713	keg	600	616	805	635	721
colony	629	630	680	526	918	latch	511	544	683	598	637
contract	578	641	694	561	747	lipstick	547	557	656	546	681
diamond	562	581	623	492	700	lottery	545	543	666	567	732
dish	554	530	621	460	646	mausoleum	812	910	792	673	941
fork	575	566	681	505	705	milestone	575	661	706	641	677

garden	603	542	611	507	624	onslaught	642	627	785	733	782
item	514	529	615	495	637	overtime	574	557	669	516	650
jungle	572	562	629	482	690	parachute	619	630	742	559	868
literature	571	645	653	601	676	pinball	591	618	673	529	670
lodge	512	595	578	613	661	quill	645	658	800	731	755
nerve	481	506	566	498	834	roster	573	571	858	700	1053
odour	579	621	725	583	715	sausage	627	623	713	513	802
pistol	505	565	627	541	744	silicon	681	725	802	612	744
pitch	538	543	661	530	641	simile	677	808	916	668	957
plastic	559	609	686	560	676	sonata	721	692	825	739	821
restaurant	543	560	638	499	668	syringe	701	674	772	590	706
ship	585	588	672	467	665	trinket	621	627	805	813	713
television	586	583	635	509	675	wig	536	522	679	574	814
temple	545	564	712	543	647	zenith	617	648	730	808	855
argument	606	603	681	548	752	asylum	704	684	821	660	757
art	536	526	597	537	696	bib	598	585	808	642	759
cane	604	581	671	501	769	coupon	619	650	767	694	709
cent	618	660	719	730	625	currant	644	665	775	623	849
chapel	634	659	701	569	703	dent	536	535	739	663	799
cheese	599	587	624	547	816	gutter	558	551	648	605	692
concert	601	633	673	537	783	icicle	709	797	789	665	755
creek	628	592	685	564	560	knuckle	598	603	669	630	727
depth	563	531	641	553	660	lint	559	536	676	548	725
disease	584	540	615	592	747	monsoon	722	657	788	700	822
entrance	594	573	648	639	707	narcotic	713	644	870	707	788
estate	674	610	665	524	671	octave	640	593	727	614	869
fund	598	609	713	610	666	payday	729	658	875	674	872
golf	512	507	577	519	772	peg	541	549	628	581	694
harbour	605	603	636	583	707	picket	596	574	799	625	762
idea	519	550	604	599	727	pizza	624	607	641	547	731
length	533	524	599	578	678	plywood	744	618	777	650	823
magnet	553	560	609	538	777	porridge	623	620	700	573	743
mercury	609	594	662	572	729	rostrum	736	553	636	666	823
moisture	578	585	609	597	866	scrapbook	773	705	813	638	680
passion	598	648	670	539	741	shotgun	627	581	734	567	722
pocket	556	573	625	575	703	shredder	664	658	854	649	837
salad	571	621	631	608	669	shutter	681	625	754	571	778
science	659	611	707	495	798	tripod	678	633	713	582	813
Statue	702	660	790	615	763	varnish	611	619	707	613	776

Table 2. Results of simultaneous regression analyses predicting mean item RT for each task from measures of phonetic features of the onset phoneme, word frequency, stimulus regularity and letter length. The summary data for each task include the overall R^2 for the complete predictor set, the β weight for each predictor variable, and the change in R^2 (ΔR^2) resulting from removing the set of phonetic predictors (Phonetic Total) and each of the other individual predictors from the regression equation (for all measures *= $p < .05$, ** $p < .01$, *** $p < .001$).

PREDICTOR	WNAM		W/NWNAM		LEXNAM		LDT		SEMCAT	
	β	ΔR^2	β	ΔR^2	β	ΔR^2	β	ΔR^2	β	ΔR^2
Voicing	-4.9		-4.6		-17.5		11.4		-2.7	
Nasal	-26.9		-28.2		-40.5		5.1		24.2	
Fricative	52.5		54.1		66.8		3.7		19.0	
Liquid	-33.9		-35.6		-37.5		-3.5		5.4	
Affricate	29.9		63.9		-12.7		-7.2		2.6	
Bilabial	1.9		2.8		-4.1		-17.8		9.5	
Labio-dental	-38.1		-34.2		-62.7		-28.6		-14.2	
Palatal	-31.2		-47.4		-16.8		-14.2		-13.3	
Velar	12.3		15.4		26.2		-15.9		-0.3	
Glottal	2.2		-7.3		-18.4		-24.1		-11.2	
Phonetic Total		.122***		.156***		.131***		.017		.001
Regularity	8.7	.017	6.7	.011	6.9	.007	1.9	.001	3.5	.000
Length	9.2	.045	10.3	.061***	1.9	.001	0.63	.000	-1.7	.001
Log Freq.	-15.4	.122***	-14.6	.120***	-25.8	.220***	-24.7	.271***	-23.1	.129***
R^2		.349***		.396***		.376***		.317*		.154

Table 3. Estimates of RT mean and variance, parameters of the best-fitting ex-Gaussian distribution (all ms units, except variance, ms^2), and percent error (misses and wrong responses) for each condition (HF = High Frequency, LF = Low Frequency, A = Animate, I = Inanimate) of the five tasks of Experiment 1.

Task	Measure	HF-A	HF-I	LF-A	LF-I	Overall
WNAM	Mean	569	569	610	629	595
	Variance	8456	8387	11723	15595	11040
	Mu	499	494	518	526	509
	Sigma	38.7	34.2	39.2	47.2	39.8
	Tau	71.1	76.1	91.9	104.1	85.8
	Miss	0.0	0.1	0.2	0.3	0.1
	Wrong	0.8	0.7	2.1	2.9	1.6
W/NWNAM	Mean	583	584	622	635	606
	Variance	10006	10133	14368	15331	12460
	Mu	509	505	528	538	520
	Sigma	43.5	39.4	43.2	54.2	45.1
	Tau	74.8	79.1	94.2	96.7	86.2
	Miss	0.0	0.2	0.0	0.2	0.1
	Wrong	0.7	0.8	2.1	2.9	1.7
LEXNAM	Mean	642	650	714	749	680
	Variance	11960	13438	20651	25774	17956
	Mu	554	554	585	604	574
	Sigma	47.9	50.1	52.5	55.9	51.6
	Tau	87.4	95.8	128.1	145.7	114.2
	Miss	0.3	0.6	6.0	7.3	3.5
	Wrong	0.4	0.6	0.8	1.1	0.7
LDT	Mean	518	535	590	620	566
	Variance	12742	14798	24522	29096	20290
	Mu	425.7	424.7	449.9	464.8	441.3
	Sigma	45.2	36.8	42.7	50.8	43.9
	Tau	91.8	110.1	140.7	154.8	124.3
	Miss	0.2	0.1	0.3	0.1	0.1
	Wrong	2.2	1.9	7.6	9.4	5.3
SEMCAT	Mean	642	748	726	804	730
	Variance	24505	28204	30398	34020	29282
	Mu	502	596	557	632	572
	Sigma	48.7	58.4	55.5	68.8	57.8
	Tau	140.6	152.0	168.8	172.2	158.4
	Miss	1.4	3.5	3.1	3.5	2.9
	Wrong	4.1	7.5	10.6	6.4	7.2

Table 4. Effect magnitudes and results of inferential testing for the NAM and W/NWNAM tasks (for all effects except the grand mean, * = .1 > p ≥ .05, ** = .05 > p ≥ .01, *** = p < .01). Effects are, Grand Mean, G, Task, T = W/NWNAM – WNAM, Frequency, F = Low – High Frequency, and Animacy, A = Inanimate – Animate. Units are ms, except Wrong, which are percentages.

Effect	Mean	Variance	μ	σ	τ	Wrong
G	600.2	11750	514.3	42.5	86.0	1.63
T	11.3	1419	11.0	*5.3	0.4	0.04
T x F	-5.3	-458	0.5	0.5	-5.9	0.01
T x A	-2.5	-1357	2.3	1.7	-5.2	0.13
T x F x A	-7.4	-3105	1.5	2.6	-9.0	-0.31
F	***47.7	***5009	***26.0	7.0	***21.5	***1.76
A	***8.3	**1223	2.0	2.6	6.0	*0.41
F x A	***15.7	2389	*13.9	*13.8	2.7	*0.85

Table 5. Treatment effects and results of inferential testing for the LDT and LEXNAM tasks (for all effects except the grand mean, * = $.1 > p \geq .05$, ** = $.05 > p \geq .01$, *** = $p < .01$). Effects are, Grand Mean, G, Task, T = LEXNAM – LDT, Frequency, F = Low – High Frequency, and Animacy, A = Inanimate – Animate. Units are ms, except for Total Errors, which are percentages.

Effect	Mean	Variance	μ	σ	τ	Total Error
G	627.2	19123	507.8	47.7	119.3	4.85
T	***123.1	-2334	***133.0	7.7	-10.1	*-1.17
T x F	6.7	-2526	8.0	-0.5	-1.5	0.20
T x A	-1.8	-15	2.0	3.0	-3.2	0.34
T x F x A	14.8	1127	2.5	-15.3	13.4	-0.87
F	***82.1	***11776	***36.2	5.5	***46.0	***6.59
A	***22.6	***3308	8.0	1.3	**14.6	**0.82
F x A	***20.2	*3082	*17.2	8.8	2.5	**1.40

Table 6. Treatment effects and results of inferential testing for the SEMCAT task (for all effects except the grand mean, * = $.1 > p \geq .05$, ** = $.05 > p \geq .01$, *** = $p < .01$). Effects are, Grand Mean, G, Frequency, F = Low – High Frequency, and Animacy, A = Inanimate – Animate. Units are ms except for Total Errors which are percentages.

Effect	Mean	Variance	μ	σ	τ	Total Errors
G	729.7	29282	571.5	57.8	158.4	10.04
F	***69.4	***5855	***45.1	**8.6	***24.2	***3.59
A	***92.0	**3660	***84.7	**11.5	7.4	0.90
F x A	***-27.3	-77	*-19.4	3.6	-8.0	***-9.36

Table 7. Task effects (T) and their interactions with animacy (A) and frequency (F) effects for the 10 possible pairwise comparisons of the five tasks. Results of inferential tests are reported as: * = $.1 > p \geq .05$, ** = $.05 > p \geq .01$, *** = $p < .01$. Task effects are scored as the first task in the comparison minus the second task eg. T (W/NWNAM vs. WNAM) = W/NWNAM – WNAM.

Pair	Effect	Mean	Variance	μ	σ	τ	% Miss	% Wrong
W/NWNAM	T	11.3	1420	10.9	*6.5	0.2	-	0.04
vs.	T×F	-5.5	-457	0.5	0.4	-6.2	-	0.01
WNAM	T×A	-2.5	-1357	2.3	1.1	-4.8	-	0.13
	T×F×A	-7.4	-3105	1.6	2.7	-9.0	-	-0.31
LEXNAM	T	***83.0	***5496	***54.5	5.1	***28.3	-	***-0.93
vs.	T×F	***40.4	***5734	**13.9	-2.1	***27.4	-	***-1.31
W/NWNAM	T×A	***14.6	*2755	5.9	-0.2	9.2	-	-0.24
	T×F×A	*15.5	2809	3.6	-14.1	10.9	-	-0.66
LEXNAM	T	***94.3	***6916	***65.4	***11.6	***28.5	-	***-0.89
vs.	T×F	***34.9	***5277	**14.4	-1.7	***21.2	-	***-1.29
WNAM	T×A	***12.1	1398	*8.1	0.9	4.4	-	-0.10
	T×F×A	8.1	-296	5.2	-11.4	1.9	-	-0.97
W/NWNAM	T	***40.1	***-7830	***78.5	2.7	***-38.2	-	***-3.63
vs.	T×F	***-33.7	***-8259	-5.8	1.7	***-28.7	-	***-4.70
LDT	T×A	***-16.4	-2770	-3.8	3.3	**-13.1	-	-0.31
	T×F×A	-0.8	-1682	-1.0	-1.3	2.4	-	-1.30
WNAM	T	*28.8	***-9250	***67.7	-3.8	***-38.4	-	***-3.67
vs.	T×F	***-28.2	***-7802	-6.3	1.3	***-22.5	-	***-4.71
LDT	T×A	***-13.9	-1413	-6.0	2.2	-8.3	-	-0.44
	T×F×A	6.6	1423	-2.6	-3.9	11.4	-	-0.99
LEXNAM	T	***123.1	-2334	***133.1	**7.8	*-9.9	***3.39	***-4.56
vs.	T×F	6.7	-2525	8.1	-0.4	-1.3	***6.20	***-6.00
LDT	T×A	-1.8	-15	2.1	3.1	-3.9	0.88	-0.54
	T×F×A	14.7	1127	2.6	*-15.3	13.3	1.09	-1.96
SEMCAT	T	***123.9	***16822	***51.7	***11.3	***72.7	-	***5.52
vs.	T×F	***24.3	1074	***18.8	1.5	5.6	-	0.96
NWNAM	T×A	***84.9	*3115	***81.6	*8.6	3.4	-	-0.88
	T×F×A	***-39.4	-913	**-34.2	-11.6	-6.2	-	***-8.32
SEMCAT	T	***135.2	***18242	***62.6	***17.8	***72.9	-	***5.56
vs.	T×F	***18.8	617	***19.3	1.9	-0.6	-	0.98
WNAM	T×A	***82.4	1758	***83.8	**9.7	-1.4	-	-0.75
	T×F×A	***-46.8	-4018	** -32.6	-8.9	-15.2	-	***-8.63

SEMCA	T	**40.9	***11326	-2.7	*6.2	***44.4	-0.67	***6.46
vs.	T×F	**16.1	**4660	4.9	3.6	***21.8	***5.37	***2.27
LEXNAM	T×A	***70.3	360	***75.7	8.8	-5.8	0.55	-0.64
	T×F×A	***54.9	-3722	**37.8	2.5	-17.1	**2.67	***7.67
SEMCA	T	***164.0	***8992	***130.2	***14.0	***34.5	***2.72	***1.89
vs.	T×F	-9.4	***7185	**13.0	3.2	***23.1	0.83	***3.73
LDT	T×A	***68.5	345	***78.8	**11.9	-9.7	**1.44	-1.19
	T×F×A	***40.2	-2595	**35.2	-12.8	-3.8	**1.58	***9.62

Table 8. Estimates of RT mean and variance, parameters of the best-fitting ex-Gaussian distribution for the Delayed Naming task of Experiment 2 (all ms units, except variance, ms^2), and percent error (misses and wrong responses) for each condition (HF = High Frequency, LF = Low Frequency, A = Animate, I = Inanimate) of Experiment 2.

Measure	HF-A	HF-I	LF-A	LF-I	Overall
Mean	400	389	402	396	397
Variance	11765	9155	14729	11708	11839
Mu	310	311	307	314	312
Sigma	40.3	47.2	40.9	48.6	44.3
Tau	90.7	72.4	95.6	82.1	85.2
Miss	1.1	1.5	1.1	1.4	1.3
Wrong	1.4	0.5	2.4	3.9	2.0

Table 9. Treatment effects and results of inferential testing for Delayed Naming task of Experiment 2 (for all effects except the grand mean, * = $.1 > p \geq .05$, ** = $.05 > p \geq .01$, *** = $p < .01$). Effects are, Grand Mean, G, Frequency, F = Low – High Frequency, and Animacy, A = Inanimate – Animate. Units are ms except for Total Errors, which are percentages.

Effect	Mean	Variance	μ	σ	τ	Total Error
G	396.7	11839	311.7	44.3	85.2	3.31
F	-4.0	-2758	3.1	-1.0	-7.3	** -2.08
A	** -8.9	-2815	7.1	*7.3	*** -15.9	0.63
F x A	4.9	-411	0.7	0.9	4.7	2.26

Figure Captions

Figure 1. Subject average vincentiles (○ symbols = animate, ● symbols = inanimate) with standard error bars and fitted ex-Gaussian vincentiles (solid lines = animate, dotted lines = Inanimate) for (a) WNAM, (b) W/NWNAM, (c) LEXNAM, (d) LDT, (e) all subjects in SEMCAT, and (f) subjects in SEMCAT with < 3% misses. Labels indicate high frequency animate (HF-A) and inanimate (HF-I) and low frequency animate (LF-A) and inanimate (LF-I) conditions.

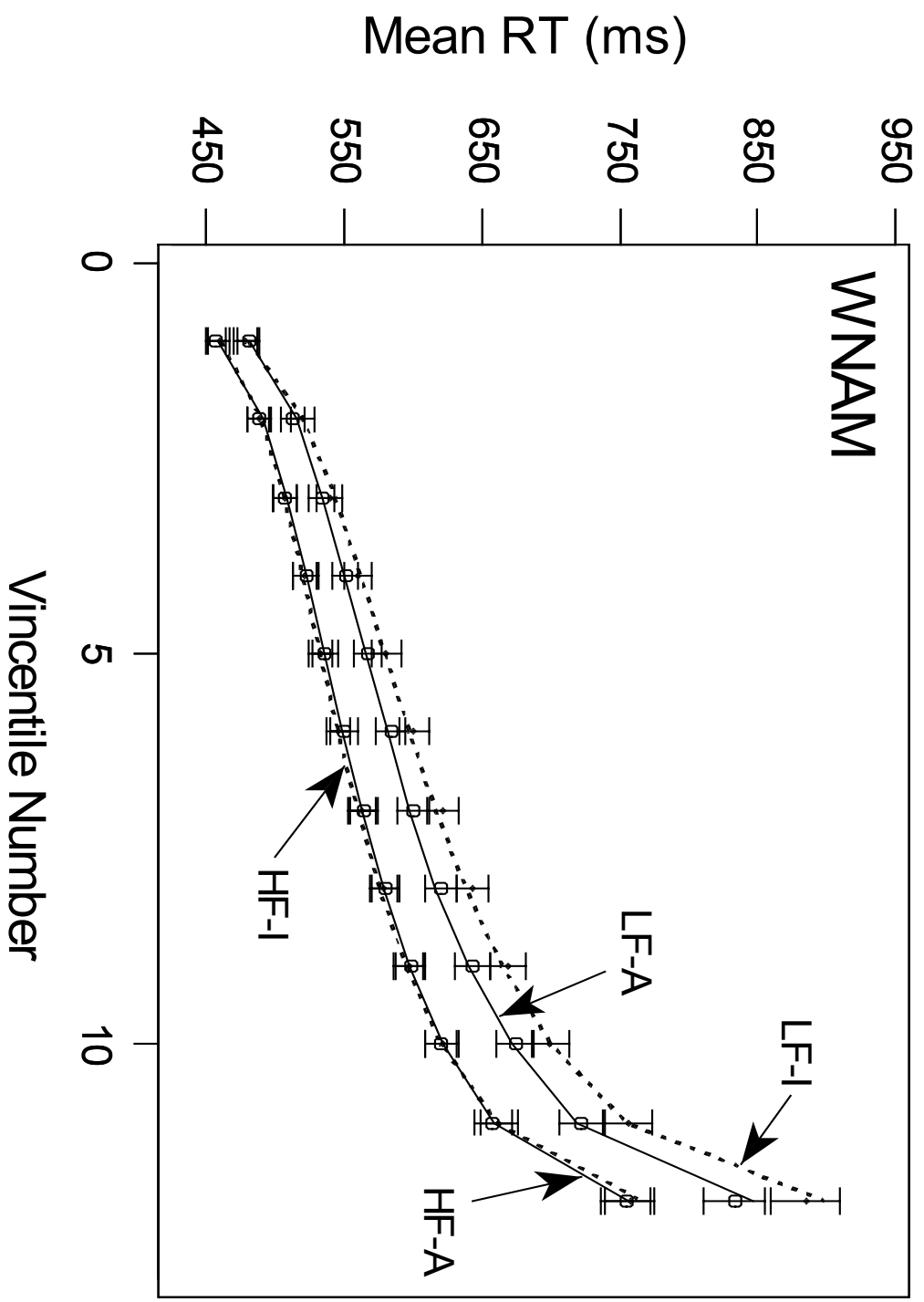
Figure 2. Plots of pairs of measures averaged over subjects: (a) Mean RT against RT Variance, (b) ex-Gaussian parameters μ against τ , and (c) Mean RT against total percent error (misses plus wrong responses). Data from different tasks are indicated by labels and by a different symbol for each task (+: WNAM, ×: W/NWNAM, ○: LEXNAM, ●: LDT, and ◆: SEMCAT). Lines show frequency effects, joining results for low frequency words (always at the left end of the line) and high frequency words (always at the right end of the line) of the same type (solid lines for inanimate words, dotted lines for animate words).

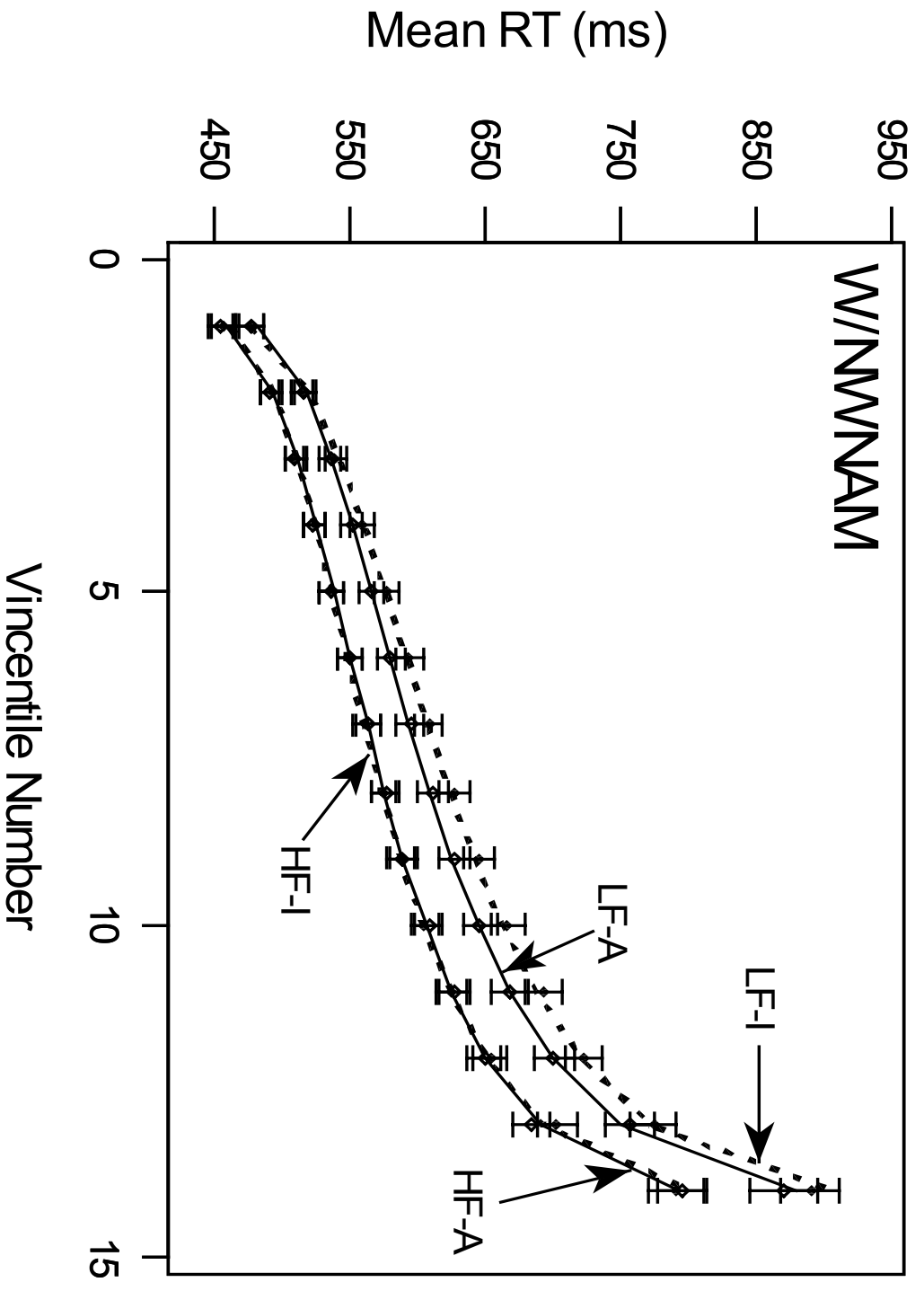
Figure 3. Vincentile values for frequency and animacy main effects and their interaction as a function of vincentile number for (a) the WNAM task and (b) the W/NWNAM Task.

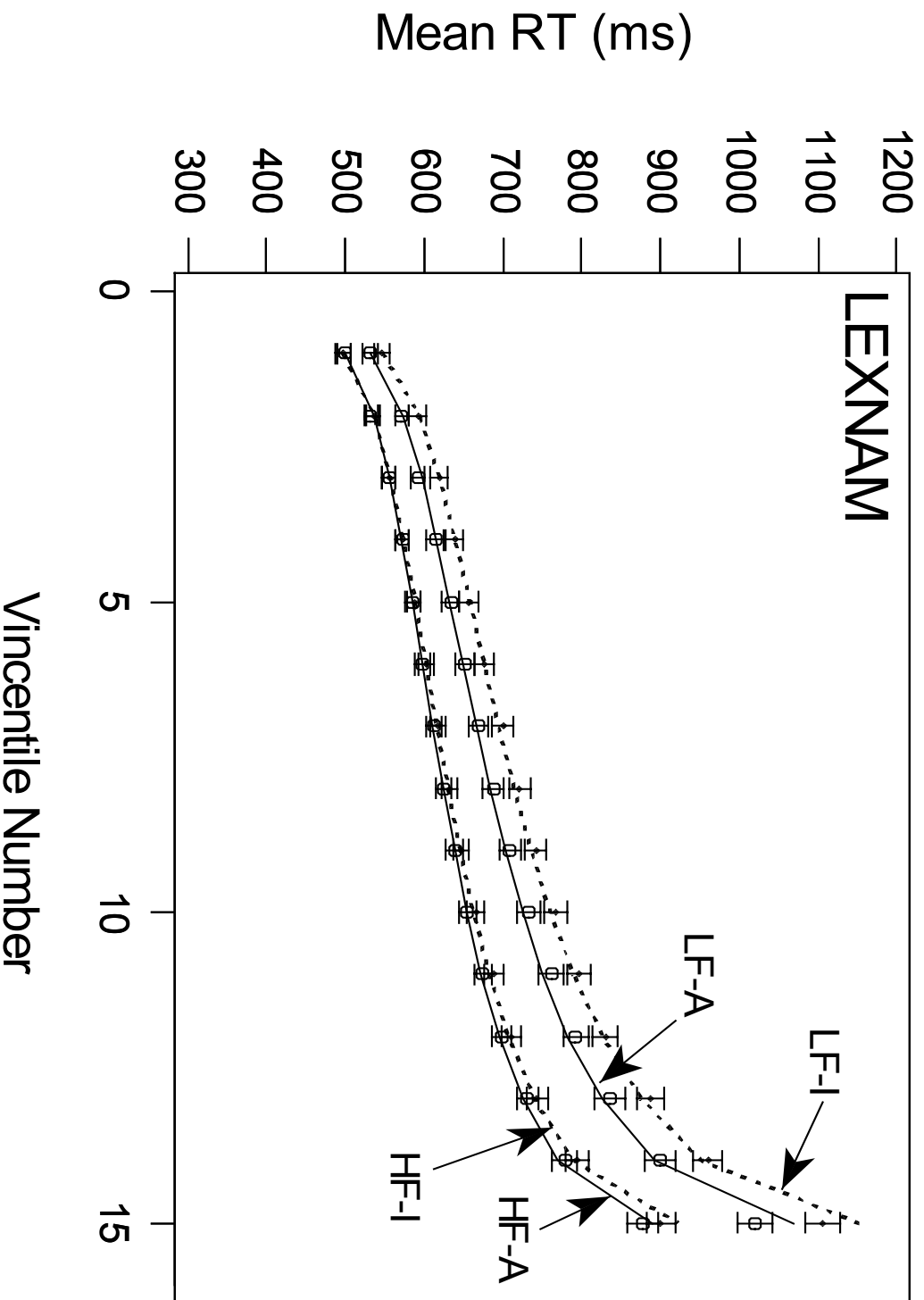
Figure 4. Vincentile values for frequency and animacy main effects and their interaction as a function of vincentile number for (a) the LEXNAM task and (b) the LDT.

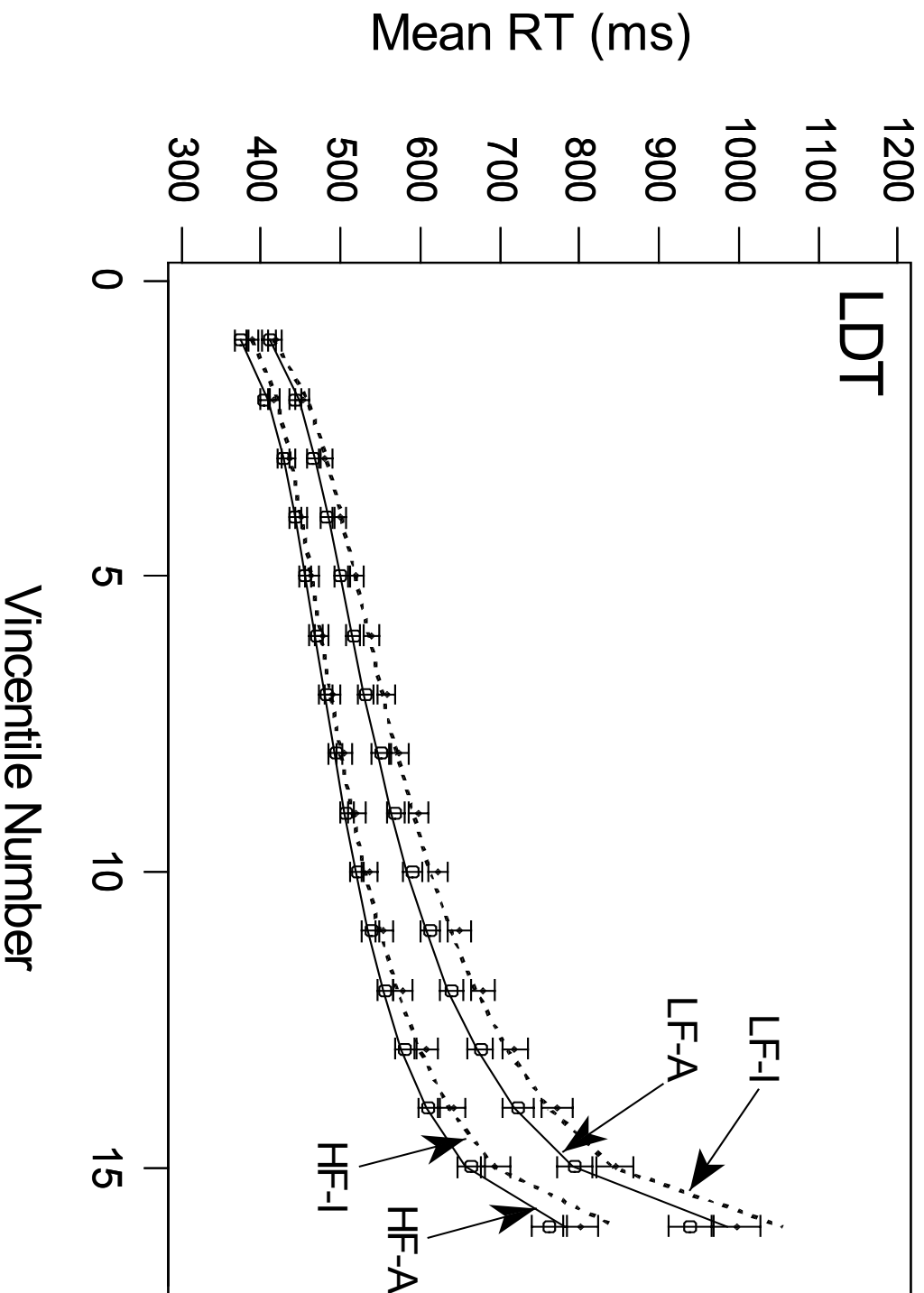
Figure 5. Vincentile values for frequency and animacy main effects and their interaction as a function of vincentile number for the SEMCAT task.

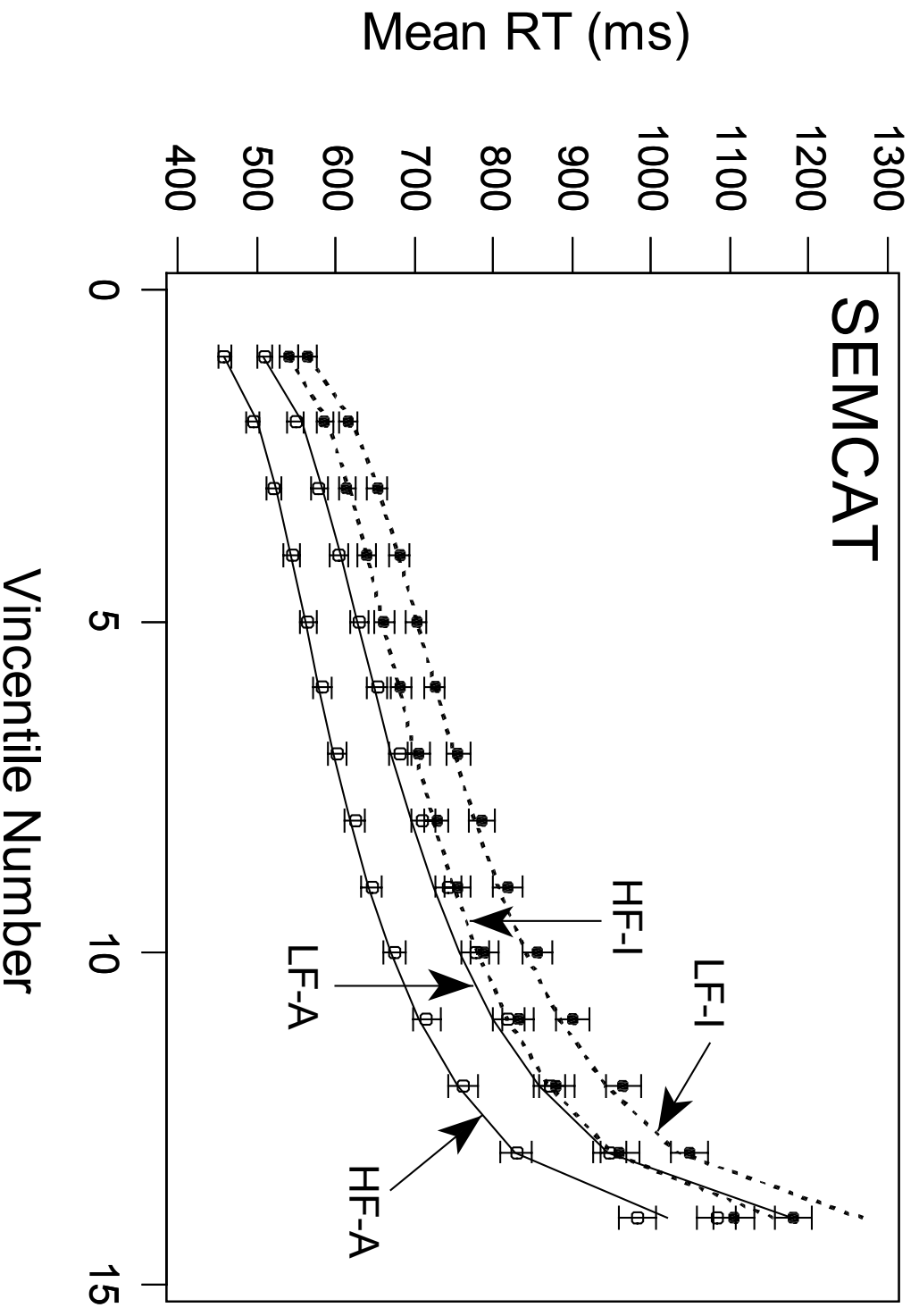
Figure 6. Subject average vincentiles (○ symbols = animate, ● symbols = inanimate) with standard error bars and fitted ex-Gaussian vincentiles (solid lines = animate, dotted lines = Inanimate) for the DELNAM task.

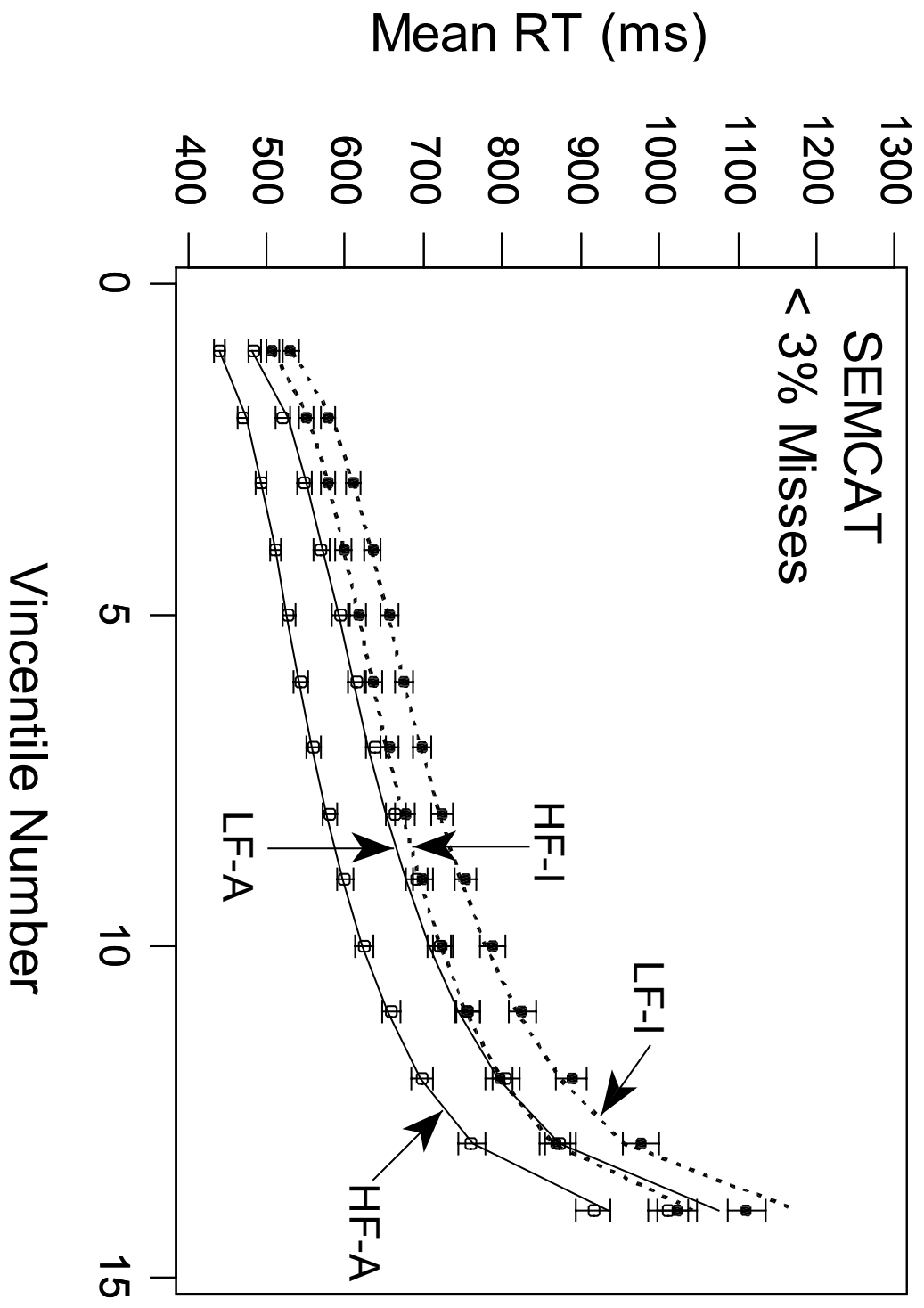


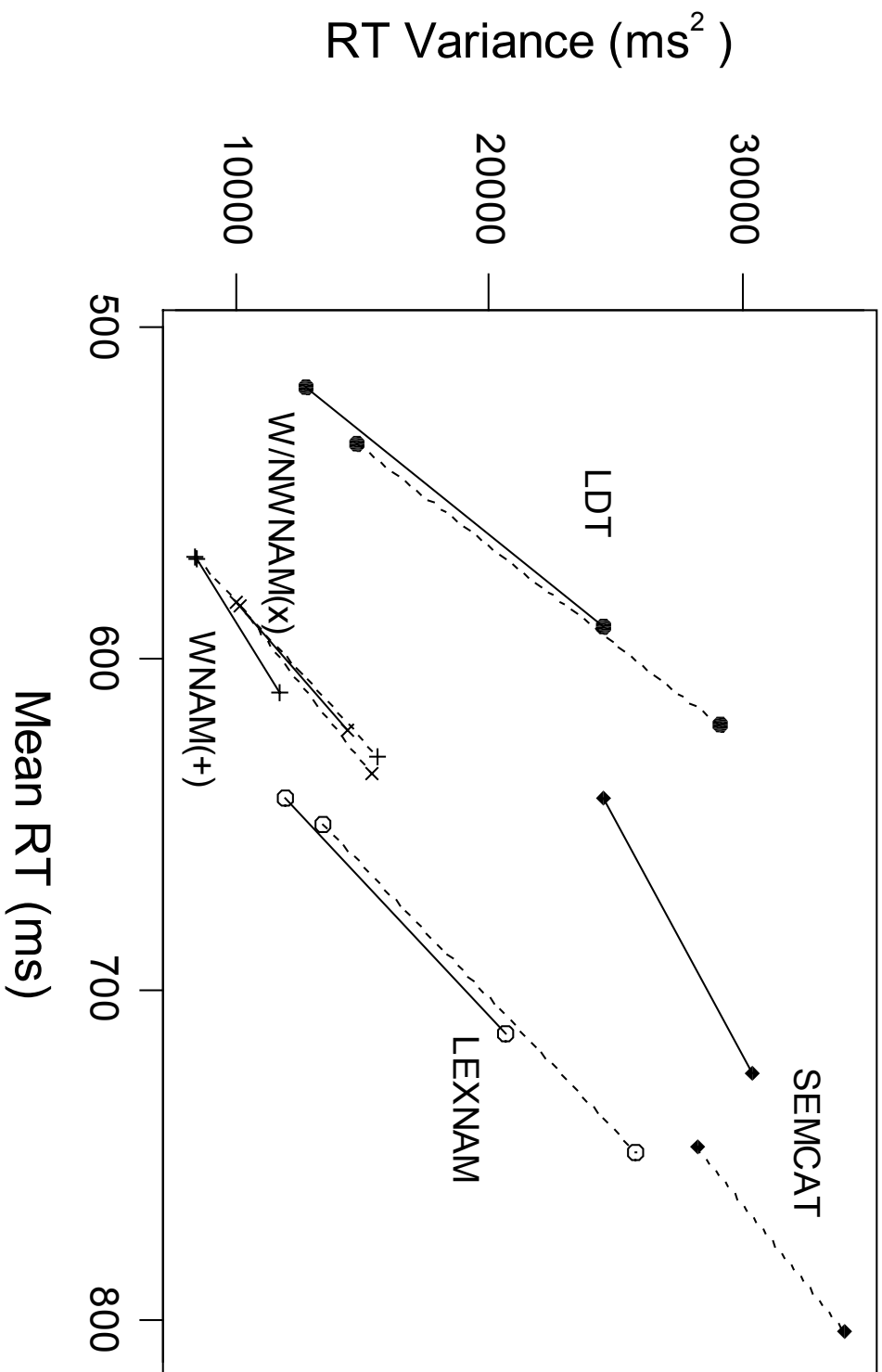


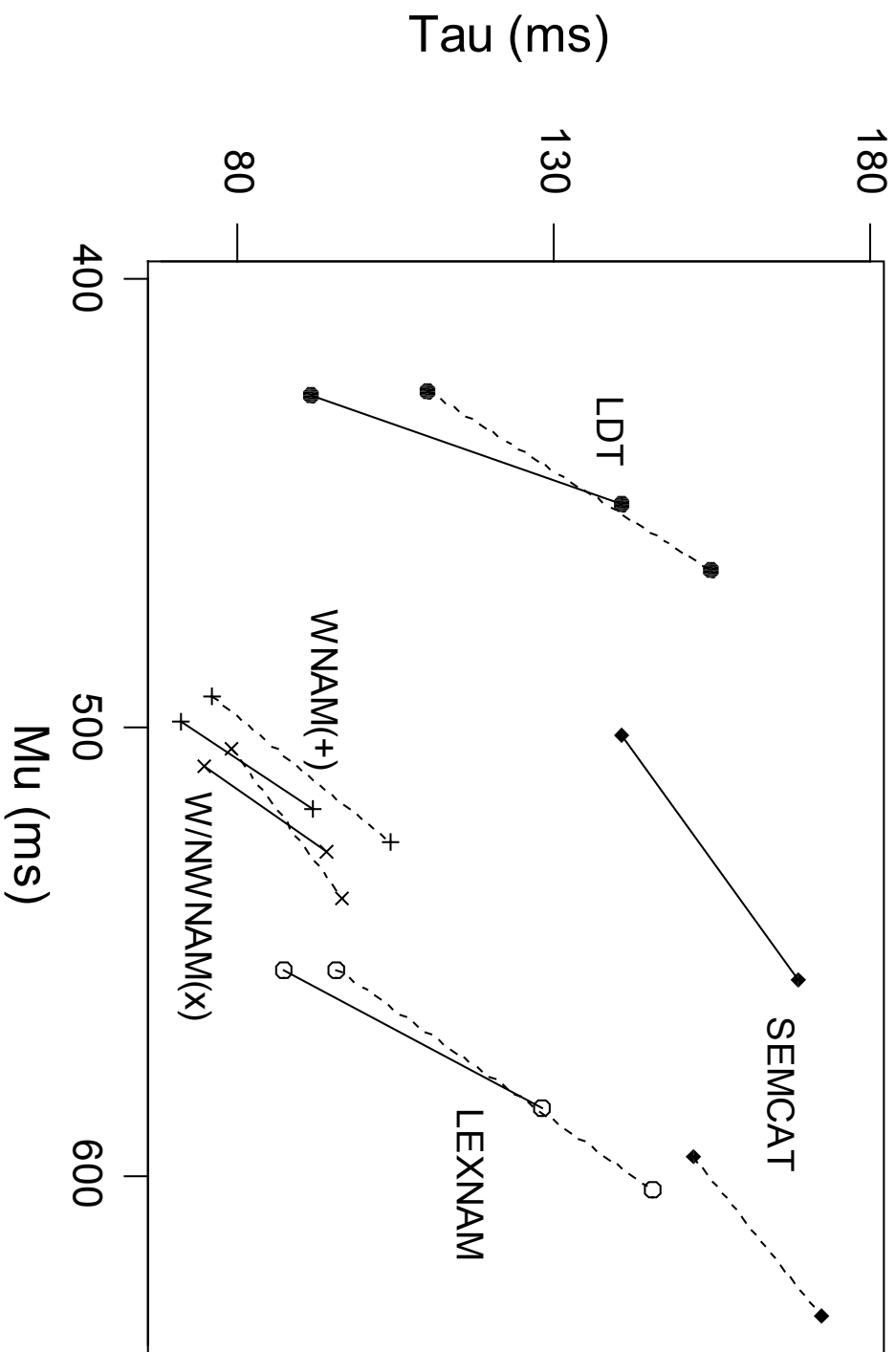


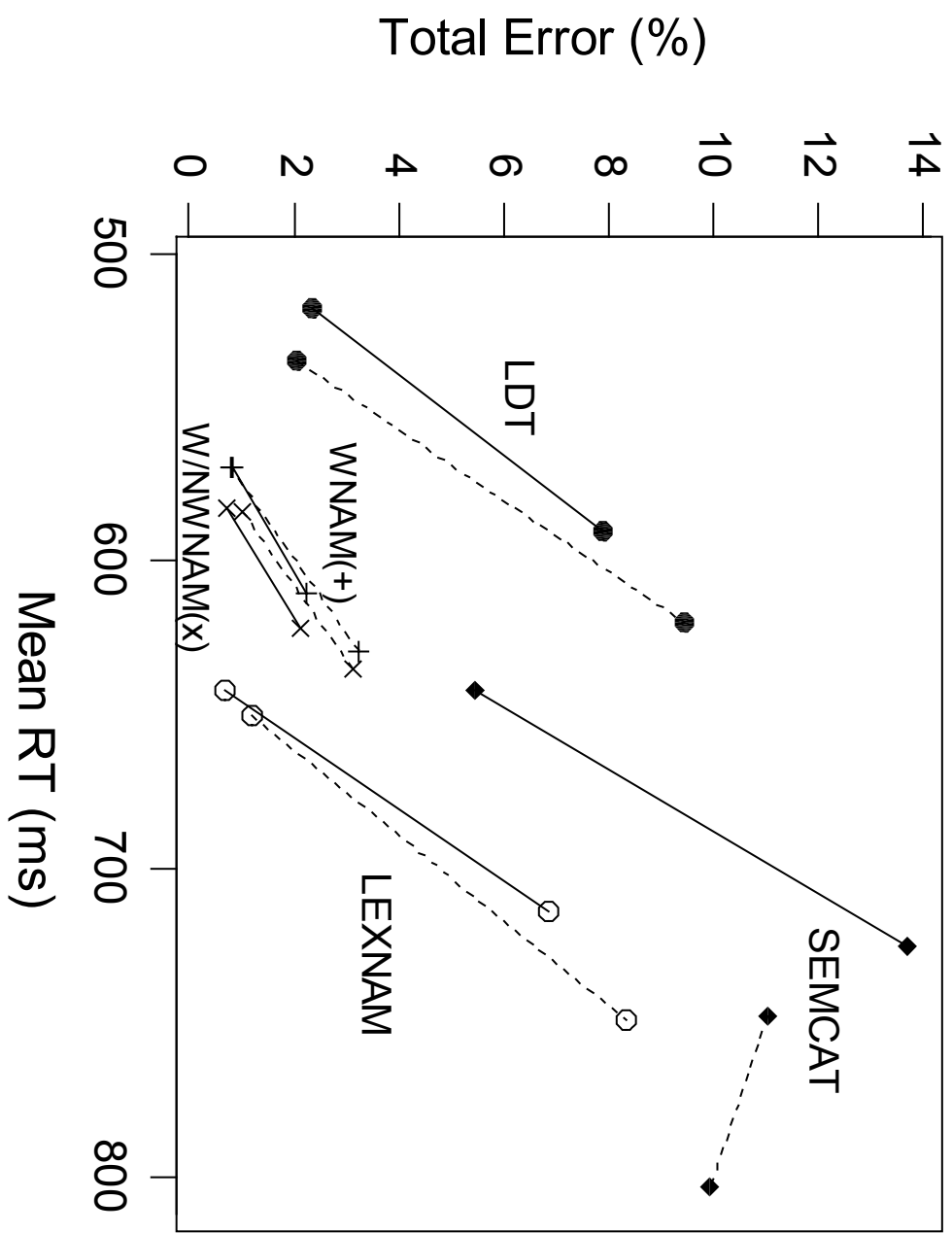


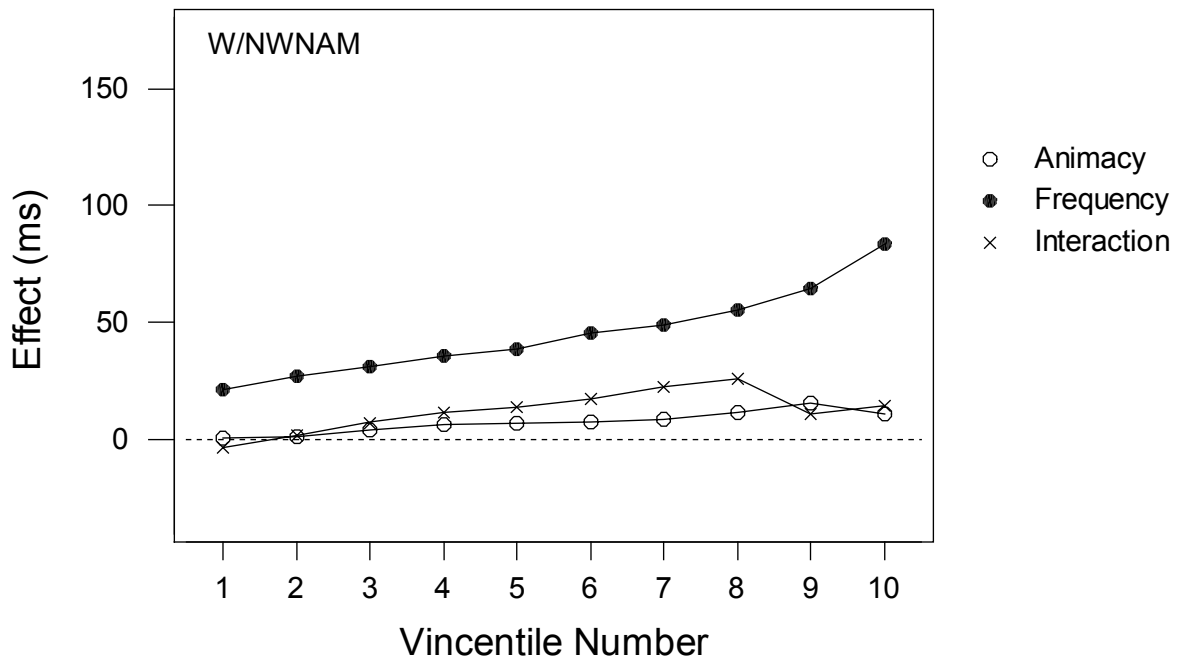
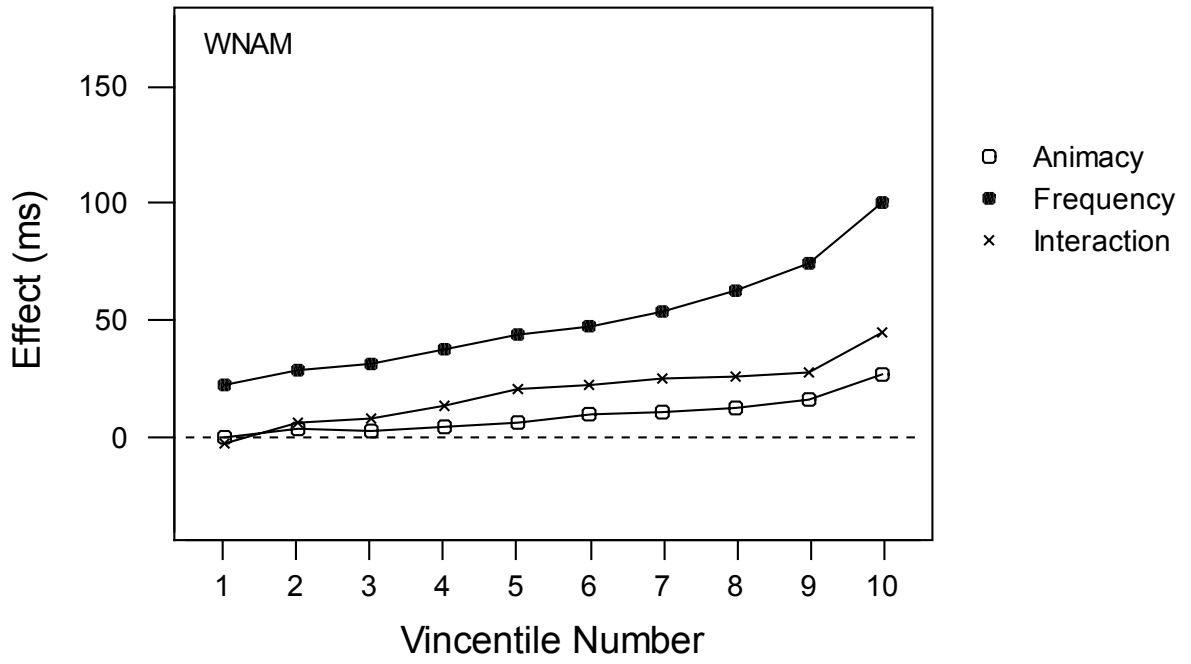


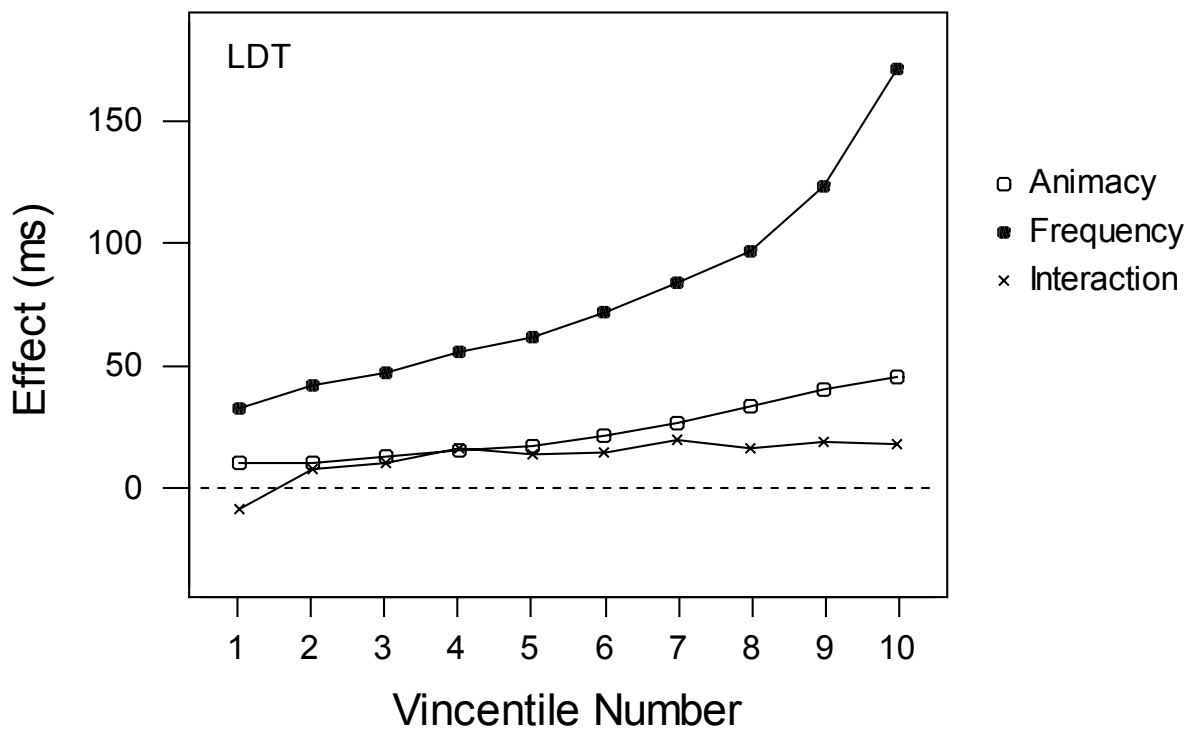
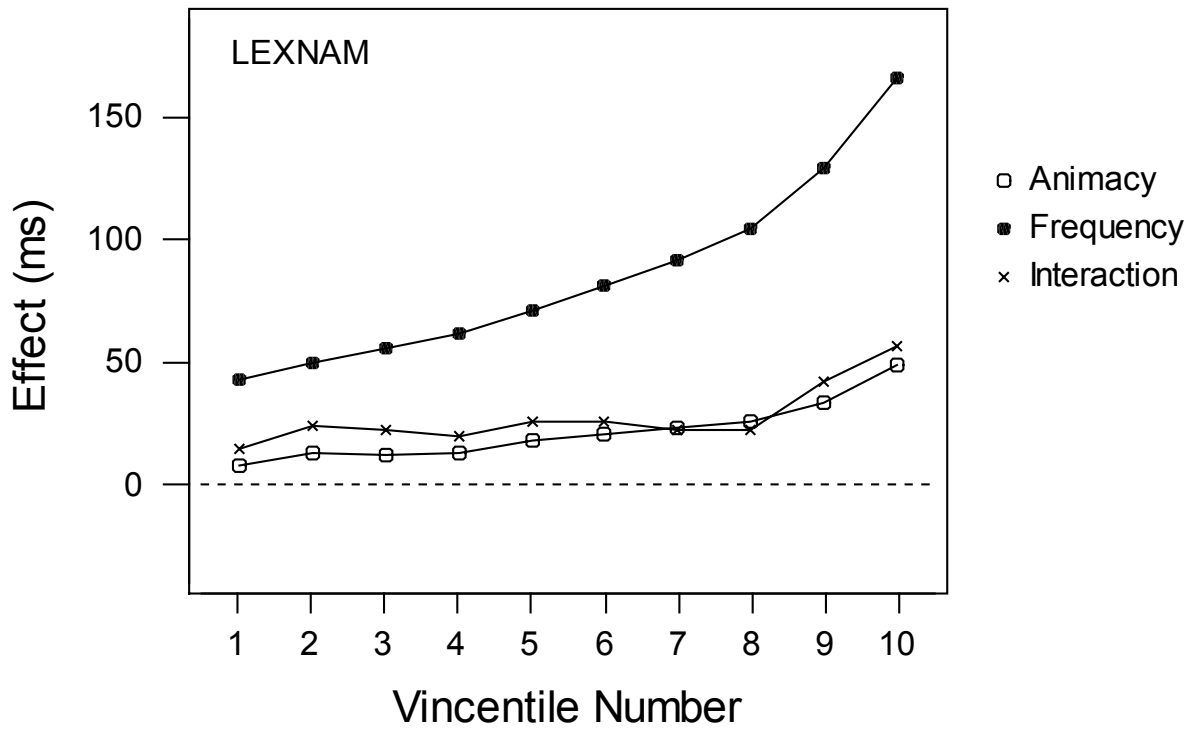


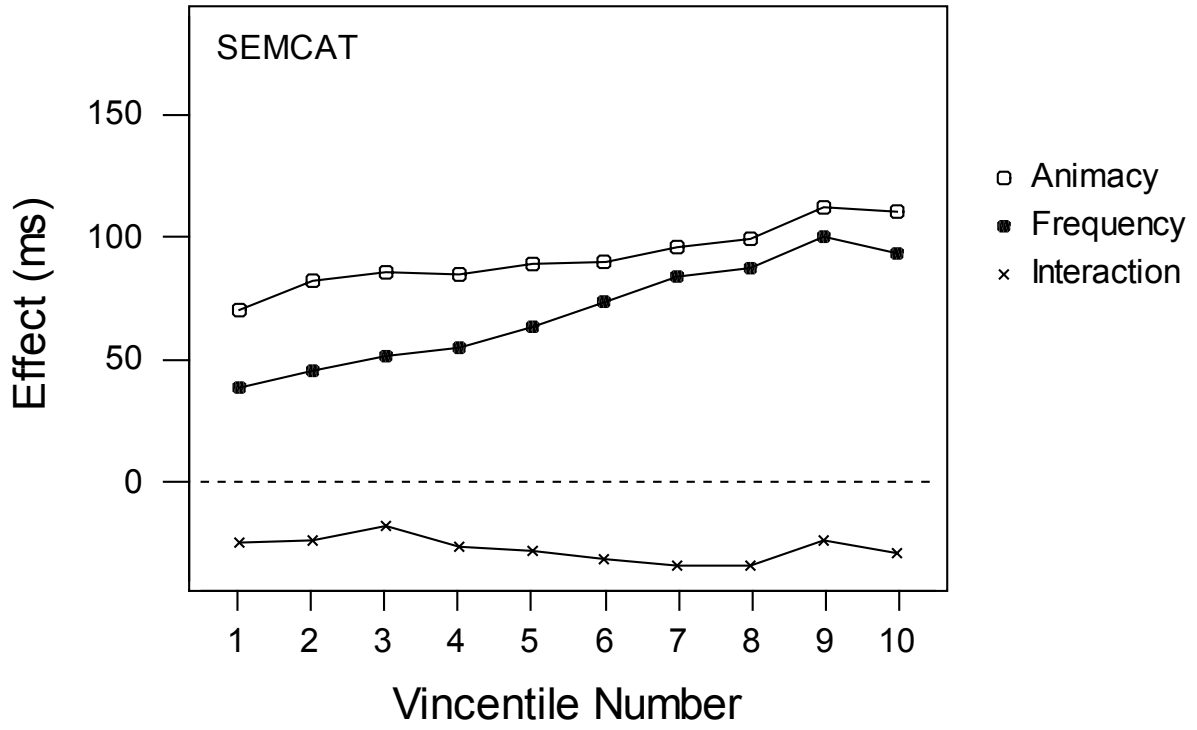


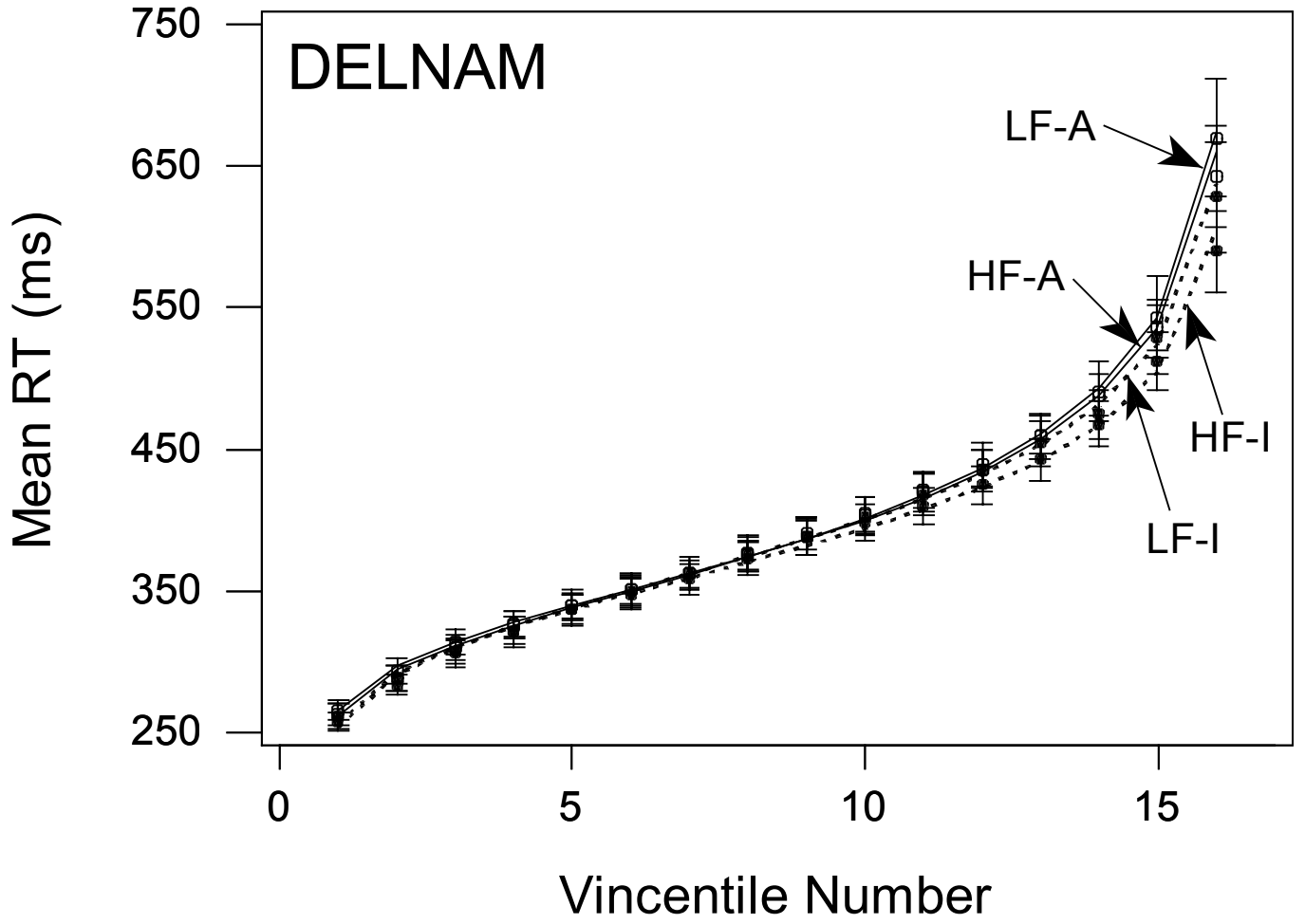












Footnotes

¹ Words with 6 or more letters have few neighbors (average $N < 1$; Andrews, 1997) so only short LF words have the potential to generate high summed lexical activity and to yield fast “word” responses.

² The first two cumulants are the mean (κ_1) and variance (κ_2), the first absolute moment and the second central moment respectively, and the third cumulant is the third central moment (κ_3). The third cumulant estimates the asymmetry or skew of the distribution. It equals zero when the distribution is symmetric (eg. a normal distribution), is negative when the distribution has a tail to the left, and is positive when the distribution has a tail to the right (eg. the exponential distribution). It is important to distinguish between theoretical cumulants with respect to a particular probability distribution, $p(x)$, (eg. $\kappa_1 = \int xp(x)dx$, $\kappa_2 = \int (x - \kappa_1)^2 p(x)dx$, and $\kappa_3 = \int (x - \kappa_1)^3 p(x)dx$) and the formula usually applied to estimate them from a sample, $x_i, i = 1..n$ (ie. $\hat{\kappa}_1 = \sum x_i / n$, $\hat{\kappa}_2 = \sum (x_i - \hat{\kappa}_1)^2 / (n - 1)$, and $\hat{\kappa}_3 = \sum (x_i - \hat{\kappa}_1)^3 / (n - 1)$). The sample formulae replace integration in the theoretical formulae with summation over the sample, allowing the frequencies of observed values to estimate the underlying probability distribution without assuming its form. It is the estimates from the sample formulae for higher cumulants such as skew that are neither robust nor efficient.

³ We have adopted the term “vincentile” rather than the term quantile adopted by Ratcliff (1979) because a quantile is a value below which a given proportion of the distribution occurs, whereas a vincentile is the average of values between pairs of quantiles. Ratcliff’s (1979, p. 449) algorithm for calculating vincentiles clearly illustrates this point. For an ordered sample of n observations write out a new list with each value repeated v times. The averages of each successive set of v values in the new list are the v vincentiles. Estimates of vincentiles and quantiles tend to be the same as the number of vincentiles approaches the sample size. In particular, quantile and vincentile estimates are identical when each interval between quantile estimates contains only one observation. Linearly interpolated quantile estimates and vincentile estimates are identical when each interval contains at most two different observations (Ratcliff, 1979, p. 449, provides an example). When intervals contain more than two different observations, however, quantile and vincentile estimates may differ.

⁴ Ratcliff (1979) proved that average quantiles provide exact average parameter estimates for the exponential and Weibull distributions. An exact result was not available for the average of ex-Gaussian quantiles, so Ratcliff provides a numerical demonstration. The number of vincentiles used in the demonstration was much smaller than the individual sample sizes, so vincentiles and quantiles were likely quite different. Since we are interested in estimating the ex-Gaussian, we also used average vincentiles, because Ratcliff’s demonstration only bears on average vincentiles not average quantiles. However, we chose the number of vincentiles to be as large as possible given our sample sizes so vincentile and quantile estimates were similar.

⁵ Note that additive factors logic dictates that effects on separate processing stages result in lack of interaction only for cumulants, not the ex-Gaussian parameters, or non-linear transformations of cumulants (eg. standard deviation or τ). Consequently, a lack of interaction in μ , σ , and τ does not necessarily support influences on separate stages. However,

when we measured τ from Plourde and Besner's graph and calculated the third cumulant ($2\tau^3$) the effects of frequency and stimulus quality still appeared additive.

⁶ We are grateful to Ken and Jonathan Forster at the University of Arizona for their continued development and maintenance of this application.

⁷ The regularity rating was made on a 4-point scale ranging from 1 (completely regular) to 4 (completely irregular) by S. A.

⁸ Given the difference between the average length of high and low frequency words, it is possible that variance attributed to frequency is actually due to length (David Balota, personal communication). To address this possibility, a second set of regression analyses were conducted on the residual RTs for each item after length was partialled out. The estimates of the unique contribution of log frequency to naming and LDT performance (.136 and .261 respectively) are very similar to the raw RT analyses reported in Table 3.

⁹ Ex-Gaussian vincentiles were determined by calculating the mean of the fitted ex-Gaussian distribution between pairs of quantiles via numerical integration. The values of each quantile were determined by line search on the numerical integral of the fitted ex-Gaussian.

¹⁰ This tendency to overestimate the slowest vincentile contrasts with Ratcliff's (1979) finding that the ex-Gaussian distribution tends to underestimate the right tail of RT distribution (e.g., Figure 4, p.455). However, Ratcliff's finding may be an artifact of the identification of vincentiles with quantiles. Plots calculated as in Figure 1 exactly fit simulated ex-Gaussian data whereas plots based on identifying vincentiles with quantiles display apparent underestimation in the right tail of simulated ex-Gaussian data. It was through constructing such plots that we realized that vincentiles and quantile values can differ appreciably in some circumstances.

¹¹ A parallel analysis of the LDT data removed only 5 participants and did not affect the fit of the data.

¹² Although each subject participated in all five tasks, the manner in which the data were stored, combined with the crossing of lists with tasks, made it difficult to use task as a repeated measures' variate. This reduces the power of the analysis of task effects but should not otherwise affect the pattern of obtained results. Mean RT, RT variance and percent error were calculated for each combination of subject, task and word type and tested by ANOVA. Parallel analyses were conducted on mean RT using items rather than subjects as the unit of analysis (Clark, 1973). All effects that were significant in the subjects-based analyses were also significant by items.

¹³ For each task, 1050 data sets were constructed by randomly sampling 74 subject's data (75 for LEXNAM) with replacement from the original set of subjects for each task. Vincent averages were calculated for each resampled set of subjects and fit with the ex-Gaussian distribution. For effects over purely within subject factors, significance was determined by counting the number of resampled sets in which the effect magnitudes was less than zero (or greater than zero if the original effect was negative). The count was divided by 1050 to determine the estimated significance level (α). Tests involving a between subjects effect were performed using t-tests based on standard errors estimated from the

standard deviations of the resampled ex-Gaussian parameter distributions. Two tailed significance tests at the 0.05 level were used in all cases.

¹⁴ Average error rates were higher in the three tasks requiring a decision (LDT, LEXNAM and SEMCAT) than in the naming tasks. Tasks with higher error rates might be argued to lead to a less accurate estimate of τ because they fail to reflect RT for items that are wrongly classified. Such items are presumably likely to be more difficult and would therefore lead to relatively slow responses if the response was correct. The RT distributions for tasks with high error rates might therefore underestimate τ . However, this problem does not confound the present cross-task comparisons because the tasks with higher error rates also produced larger τ estimates (see Figure 2). The elimination of 5- 10% of slower responses from the three decision tasks compared to about 2% of naming responses may mean that the τ differences between tasks are slightly underestimated, but would not change the overall pattern of cross-task differences in τ .

¹⁵ The single exception was HF inanimate words in LEXNAM, for which correct and error responses were of equivalent speed (658 ms and 660 ms respectively) but since the error rate was only 0.6% this estimate is unreliable.