Attentional Control and Flexible Lexical Processing:
Explorations of the Magic Moment of Word Recognition

David A. Balota & Melvin J. Yap
Department of Psychology
Washington University

Address Correspondence to:  David A. Balota
Department of Psychology
Washington University
St. Louis, MO  63130
dbalota@artsci.wustl.edu
In studying any behavior, one needs to first operationalize the targeted behavior. In the present case, the targeted behavior is visual word recognition. At one level, this behavior is deceptively simple to understand. Adult readers have intimate knowledge of this acquired skill. However, as shown in this volume, the processes involved in visual word recognition are remarkably complex. This should not be surprising because experimental psychologists have been working on this topic for well over 100 years (Cattell, 1890). We believe that at least some of the complexities and controversies in this area are due to a lack of agreement in the theoretical assumptions regarding the targeted behavior, the tasks, and the analytic methods used to measure the processes underlying this behavior. In this chapter, our mission is to attempt to elucidate some of these theoretical assumptions and the methodological approaches to studying word recognition. We will use the influence of meaning on word recognition to guide our discussion, because the theoretical and methodological assumptions are nicely unearthed by considering meaning level influences.

I. MEASURING THE MAGIC MOMENT IN WORD RECOGNITION

A reasonable, yet often implicit, assumption underlying models of visual word recognition is that there is a magic moment in word processing (Balota, 1990), a discrete instant when a reader recognizes a word, but does not yet know its meaning. At first glance, this seems quite reasonable and inherent in most models of pattern recognition, i.e., how could one interpret something unless one has first recognized what that something is? In more technical terms, the magic moment is that instant when lexical identification takes place, i.e., a lexical representation is sufficiently activated for a response to be executed (Balota, Ferraro, & Connor, 1991). This event unlocks access to
meaning. For example, in activation-class models (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; McClelland & Rumelhart, 1981; Morton, 1969), the magic moment is when the activation level for a word detector exceeds some threshold, and lexical identification takes place. In search-class models (Becker, 1980; Forster, 1976; Paap, Newsome, McDonald, & Schvaneveldt, 1982), the magic moment is when there is sufficient overlap between a word’s sensory representation and its internal orthographic representation, resulting in successful search/lexical identification.

Assuming there is such a magic moment, it is critical that one can measure the processes leading up to this point. Although there are many measures of the magic moment (such as on-line reading measures and perceptual identification), this construct is most typically defined via two major tasks: lexical decision and speeded naming. In lexical decision, participants are presented with either a real word or a non-word, and the time it takes to make word/nonword binary responses is measured. In speeded naming, participants are presented with words, and the time it takes for participants to initiate a vocal response is measured. For both tasks, response latencies are often assumed to reflect word recognition processes that presumably are decoupled from meaning access, and appear to be relatively immune to attentional control mechanisms (see Coltheart et al., 2001; Murray & Forster, 2004).

The notion that lexical decision and naming latencies tap presemantic aspects of word recognition is however inconsistent with the empirical observation that semantic effects have been reliably observed in isolated lexical decision, and to a lesser extent, naming. To the extent that these two tasks are windows into the magic moment of word recognition, such a disjunction is indeed puzzling. However, it is increasingly clear that
neither lexical decision nor speeded naming reflect a magic moment. There is a basic problem with both paradigms; they measure both word identification processes as well as operations that are specific to each task.

II. Task-Appropriate Processing and the Flexible Lexical Processor

We believe that it is critical to consider word recognition within a Task-Appropriate Processing Framework. This perspective has considerable similarity to the Transfer-Appropriate Processing Approach (Morris, Bransford, & Franks, 1977; Roediger, 1990), which has been particularly useful for understanding the effects of variables in the memory domain. Memory researchers have recognized that the influence of a variable strongly depends on the tasks used to tap that variable. For example, although pictures are better remembered than words in a free recall task, words show more priming than pictures in a word-fragment completion task (Weldon & Roediger, 1987). Blaxton (1989) has demonstrated that, compared to perceptually driven tasks (e.g., word fragment completion and graphemic cued recall), conceptually driven encoding tasks (e.g., free recall, semantic cued recall, and a general knowledge test) produce better performance on conceptually driven retrieval tasks, whereas, the opposite pattern is obtained for perceptually driven retrieval tasks. The assumption underlying the transfer-appropriate processing perspective is that different memory tasks require different retrieval operations and therefore benefit from different types of processing during encoding. To the extent that there is a match between encoding and retrieval operations, performance is facilitated.

Of course, in language processing, there is a rich tradition of considering modular/dedicated systems that are immune to top-down control systems (Fodor, 1983).
We believe that this curse of automaticity has produced a relatively inflexible lexical processor that emphasizes pre-existing structures and processes. Here we argue that word recognition tasks, like memory tasks, are modulated via attentional control systems and the processes that are relevant for accomplishing the goals of the task. Just as Jacoby (1991) has argued there are no process-pure measures of memory, we would argue there are no process-pure measures of word recognition. Hence, lexical decisions will be modulated more by variables that help participants discriminate familiar words from unfamiliar nonwords, whereas naming performance benefits from manipulations that emphasize the pathways necessary for generating a correct pronunciation as quickly as possible. Armed with the task-appropriate processing perspective and the notion of a flexible lexical processor, we shall now turn to the search for a magic moment in word recognition tasks.

III. BEYOND MEASURES OF CENTRAL TENDENCY IN MEASURING THE MAGIC MOMENT IN WORD RECOGNITION

Thus far, we have discussed how task-specific characteristics may stymie efforts to measure a simple magic moment. Of course, current mental chronometric methods also make assumptions that may diminish the rate of knowledge accumulation about the underlying processes involved in measuring the magic moment. Consider, for example, the standard experimental paradigm wherein participants are presented with a set of Type A words and a set of Type B words, e.g., high and low-frequency words. If one obtains an effect of frequency on some estimate of central tendency (most typically means) in one of the windows (lexical decision or naming) into the magic moment, then one assumes that they have captured an effect of this variable on the word identification
processes. Of course, using estimates of central tendency to isolate the influence of processes extends far beyond visual word recognition and persists across virtually all domains of cognitive psychology.

However, there are at least two aspects of these chronometric assumptions that one may question. First, there is the notion that estimates of central tendency are the best way to measure the influence of a variable on performance on a response time task. For example, by using the mean as an estimate of the central tendency, one is making the implicit assumption that a variable is shifting the underlying response latency distribution, i.e., the distribution of low-frequency words is simply slower by some constant compared to the distribution of high-frequency words. However, the mean of a response time distribution can be influenced by shifting the distribution and/or skewing the distribution. Interestingly, there is even some evidence of tradeoffs between the two components, such that there is no effect in means. For example, the congruent condition in Stroop color naming, relative to the baseline, simultaneously decreases the modal portion of the response time distribution while increases skew. This finding was initially reported by Heathcote, Popiel and Mewhort (1991) and has since been replicated by Spieler, Balota and Faust (1996). In this case, there are two distinct effects in the data that are being masked by measures of central tendency because of opposing influences on the means.

There is a rich literature on measuring the influence of a variable on the shape of underlying response time distributions (see, for example, Luce, 1986). One procedure is to fit an empirical response latency distribution to a theoretical distribution such as the Weibull or ex-Gaussian distribution. With such procedures, one can measure if a variable is shifting the distribution (as typically assumed) and/or changing the shape of the
distribution (see, for example Andrews & Heathcote, 2001; Balota & Spieler, 1999;
Spieler, Balota, & Faust, 2000).

An alternative procedure for measuring the locus of an effect is to simply plot the
data as a function of Vincentiles (Andrews & Heathcote, 2001; Vincent, 1912). In this
procedure, for each participant, one rank orders the response latencies, and then produces
a mean for each bin of a given size, e.g., every 10% of the data. One can then examine
the locus of an effect of a variable, i.e., is the effect of the variable constant across all
Vincentiles or is the effect more localized at the early or late Vincentiles?

For illustrative purposes, we present the data from a recent study collected in our
laboratory, which examines the interactive effects of semantic relatedness and visual
degradation in a lexical decision task. In this study, visual degradation (clear vs. degraded)
and semantic relatedness were factorially manipulated to create four conditions, with 75
observations in each condition. As one typically finds in such studies (see, for example,
Becker & Killion, 1977), the semantic priming effect was larger for degraded items
(Related = 630 ms; Unrelated = 696 ms) than for clear items (Related = 545 ms;
Unrelated = 578 ms). Figure 1 plots the semantic priming effect across each of the
Vincentiles for the clear and degraded conditions.

As shown in this figure, the effect of semantic priming in the clear condition is
constant across the Vincentiles (which is consistent with the notion that priming simply
shifts the response time distribution), whereas the interaction between relatedness and
degradation primarily occurs at the later Vincentiles (suggesting that the interaction is
due to a disproportionate skewing of the unrelated degraded distributions). This was also
reflected in parameters from the ex-Gaussian analyses, wherein the interaction between
degradation and relatedness was mediated more by the Tau parameter (reflecting distributional skewing) than by the Mu parameter (reflecting distributional shifting).

Models such as the multiple read-out model (Grainger & Jacob, 1996) and Ratcliff, Gomez, and McKoon’s (2004) recent model of lexical decision performance are ideally suited for testing the influence of variables on underlying response latency distributions. It will also be particularly illuminating to test other models, such as the PDP model of lexical decision (Plaut, 1997) and the dual-route cascaded model of word recognition (Coltheart et al., 2001), at the level of response time distributions. Andrews and Heathcote’s chapter (this volume) provides another example of using such analyses to provide further insights into the manner in which a variable influences the processing system. The important point here is that with the increasing sophistication of the extant models, it is time to increase the sophistication of the analytical tools used to test the models.

Figure 1. Vincentile means of the participant’s lexical decision response times as a function of semantic relatedness and visual degradation.
Another assumption that most models of mental chronometry make is that the influence of a variable is up until the decision point wherein participants initiate a relatively dumb ballistic response (see Logan & Cowan, 1984, for a discussion of the point of no return). This latter ballistic response is simply an additive constant above and beyond the interesting processes that ended at the decision point. Hence, one can use the triggering of an electronic microswitch as an accurate terminal marker of the processes leading to the initiation of the response, i.e., the magic moment. However, even this assumption has qualifications. Although the triggering of a microswitch measures important temporal requirements of targeted processes, they only tap a single point in the information processing stream (Balota & Abrams, 1995), and do not capture mental operations that may operate after response onset. For example, there is evidence that lexical influences persist after response initiation (Abrams & Balota, 1991; Balota & Abrams, 1995). In these studies, participants made lexical decisions by moving a joystick handle rapidly to the right for words and to the left for nonwords. Hence, one can measure both the initiation of the response and the dynamics (e.g., acceleration, peak force, duration) of the response after the response has been initiated. One would argue that the initiation of the response is related to the magic moment, and hence variables that influence the processes tied to recognition processes should terminate at the onset of the response. However, the results from these studies indicated that word-frequency influenced the force of the response after response initiation. Hence, frequency effects extend beyond the decision point in lexical decision, and so the magic moment is not so clearly discernable. Similarly, in speeded naming, word frequency has been shown to influence both the onset and the production duration of pronunciation responses (see
Balota & Abrams, 1995). Clearly, the kinematic aspects of responses after the response has been initiated can be influenced by a lexical-level variable like word frequency. Classic models of word recognition assume that variables (e.g., word frequency) influence word identification processes up to the point when a response is initiated (i.e., the magic moment); this assumption is simply inconsistent with the post-initiation effects observed.

It is also interesting to consider how binary decision processes such as lexical decision may be implemented in the neural hardware. Are decision responses driven by relatively few neurons or a large ensemble of neurons? Is there a common neural decision system for lexical decision and semantic classification, or do different binary tasks engage different neural ensembles for each type of binary decision? Answers to such questions could provide some insight into the appropriate model. Interestingly, in a recent review that examined the neural correlates of decision making, Schall (2003) found support for at least two neural processes that drive simple decision processes. Neurons in sensorimotor structures accumulate evidence via a diffusion process (Smith & Ratcliff, 2004), while other neurons prepare and initiate overt movements. Importantly, these two processes can be dissociated, supporting the view that evidence accumulation and response production are formally dissociable at the neuroanatomical level. Coupled with the preceding discussion on post-initiation frequency effects, this suggests that variables such as word frequency could actually be influencing both the decision-making stage and the response stage at the neural level. While it is as yet unclear how one might implement such effects, models of word identification will ultimately need to consider how the brain engages such decision processes. At the very least, it will be useful for cognitive
scientists interested in models of word recognition to consider the emerging evidence regarding how the brain actually enables a response.

The observations reviewed in this section highlight how some of the methodological assumptions regarding the locus of an effect of a variable on response latencies appear to be limited. These assumptions do not simply constrain the inferences drawn from studies of word recognition, but also apply to chronometric studies in other domains of psychology. The results reviewed in this section question the basic assumption of chronometric studies; that there is a point in time when a lexical representation’s threshold is reached, at which point a response is executed. In fact, there may not be any discrete moment when word identification takes place. Rather, lexical processing is more likely to be a continuous, cascadic flow of information, where experimental variables can influence early identification, decision, and late post-decision processes (McClelland, 1979), depending upon the goals of the task (i.e., the task-appropriate processes).

IV. MEANING-LEVEL INFLUENCES IN ISOLATED WORD RECOGNITION: SOME OF THE INITIAL EVIDENCE, AND POSSIBLE CONCERNS

Armed with the above considerations and the potential limitations in any measure of lexical identification, we shall now turn to the literature concerning meaning-level influences on lexical identification processes. Again, the critical question here is whether meaning provides a top-down influence during word recognition (Coltheart et al., 2001; Harm & Seidenberg, 2004), or whether word recognition mandatorily precedes access to meaning (Becker, 1980; Morton, 1969). Here, meaning-level characteristics are usually operationalized by variables such as concreteness (the degree to which a word refers to
an object, material or person), *imageability* (the degree to which a word generates mental imagery, Cortese & Fugett, 2004), and *meaningfulness* (the degree to which a word evokes associates and other words, Toglia & Battig, 1978). However, as described later, more recent notions of semantic memory have indicated that additional measures such as network *connectivity* (how interconnected a word is to other words, Steyvers & Tenenbaum, 2005) may also be useful to consider.

Although there has been considerable interest in this topic, the empirical evidence for meaning-level influences in isolated word recognition remains relatively sparse and controversial, particularly for speeded naming. For example, Strain, Patterson, and Seidenberg (1995) observed a three-way interaction between word frequency, spelling-to-sound consistency, and imageability in speeded naming performance. Specifically, imageability effects in speeded naming were strongest for low frequency, inconsistent items, i.e., items that are difficult to name. They interpreted this finding as consistent with a tripartite model of word recognition, wherein semantic information provides more support for items that have relatively weak spelling-to-sound mapping (Strain, Patterson, & Seidenberg, 2002). Monaghan and Ellis (2002) disputed this finding, and demonstrated that the critical interaction was not significant once age of acquisition (AoA), a confounding variable, was controlled for (but see Strain et al. 2002, for a reply). Of course, Gernsbacher (1984) pointed out over twenty years ago that many word recognition studies investigating meaning-level variables have not adequately controlled for stimulus familiarity, a variable that is confounded with many semantic variables.

One possible reason for these controversies is that studies usually manipulate variables using a relatively small set of items (typically fewer than 20 items per
condition). Factorial experiments, while valuable may also produce some problems. First, Forster (2000) found that researchers in word recognition were quite good at predicting which of two words would produce faster response latencies in a lexical decision task, even though obvious variables such as frequency and length were equated. Hence, word recognition researchers appear to have implicit knowledge concerning the effects of lexical variables, and this could contaminate item selection processes. Specifically, researchers could consciously or unconsciously select items for categorical manipulations that support their hypotheses (Rosenthal, 1995), a phenomenon we refer to as Forster Fibbing. Second, lists contexts (i.e., characteristics of words in a list) vary across studies, and the varying list environments may modulate the effect of interest (Andrews, 1997; Glanzer & Ehrenreich, 1979; Zevin & Balota, 2000). For example, Glanzer and Ehrenreich (1979) found large modulations of the word-frequency effect depending on the frequency of filler words. Third, standard factorial designs categorize continuous variables, reducing power and reliability (Cohen, 1983). For example, studies that dichotomize imageability, a continuous variable, are less likely to find significant imageability effect. Finally, given the large number of variables known to influence word recognition, it is becoming virtually impossible to control for all potentially extraneous and confounding variables (Cutler, 1981). The presence of putatively contaminating factors makes it particularly vexing for the field to establish the reliability (or the lack thereof) of factors (e.g., consider the debate between Ellis & Monaghan, 2002, and Strain et al., 2002).
V. FURTHER EXPLORATIONS OF SEMANTIC EFFECTS IN VISUAL WORD RECOGNITION: 
RESULTS FROM MULTIPLE REGRESSION ANALYSES OF LARGE-SCALE DATABASES

An approach that minimizes list selection and list context effects is to examine lexical decision and naming performance for a large corpus of stimuli. There have been a number of attempts to explore the utility of this approach in the literature (e.g., Seidenberg & Plaut, 1998; Spieler & Balota, 1997; Treiman, Mullennix, Bijeljac-Babic, & Richmond-Welty, 1995). Recently, we (Balota et al., 2004) reported the results from a megastudy of lexical decision latencies (n=30) and naming latencies (n=31) for virtually all the monosyllabic words in the Kučera and Francis (1967) norms. Using multiple regression, it is possible to estimate the unique variance accounted for by targeted semantic variables, after potential confounding variables have been partialed out in earlier steps. Multiple regression also allows the language, rather than the experimenter, to define the stimulus set.

In the 3-step hierarchical regression analysis we carried out, phonological onset variables (voicing, location, and manner of articulation of word-initial phonemes) were entered in Step 1, and lexical variables (length, orthographic neighborhood density, objective and subjective frequency, feedforward and feedback consistency) were entered in Step 2. As expected, the phonological onset variables predicted significantly more variance in naming ($R^2 = .35$) than in lexical decision ($R^2 = .01$), which is consistent with the idea that naming performance is influenced by word-initial phonemes (Kawamoto & Kello, 1999). By entering the lexical variables in Step 2, we are able to determine if semantic effects exert unique effects in word recognition after other potentially confounding variables are controlled for (c.f., Gernsbacher, 1984). Here we will focus on
the semantic effects in Step 3, but we will return to the results from Step 2 in a later section.

Two parallel analyses of semantic effects were conducted. In the first analysis, we examined how well imageability predicted lexical decision and naming latencies, using a new set of imageability norms for all monosyllabic words (Cortese & Fugett, 2004). Imageability was reliably facilitatory in both tasks, with stronger effects in lexical decision ($\beta = -.27, p < .001$) than in speeded naming ($\beta = -.04, p < .05$). Motivated by the work of Strain and colleagues (1995), we also explored interactions (both two-way and three-way) between imageability, word frequency, and consistency. Unfortunately, none of the interactions were reliable ($ps > .2$), suggesting that the interaction between meaning-level variables and other variables is relatively modest when all single-syllable words are considered.

In the second analysis, we used two semantic connectivity measures based on Steyvers and Tenenbaum’s (2005) work on semantic network structures. These researchers were motivated by recent evidence suggesting that naturally occurring networks (such as the *C. elegans*, the power grid of the Western United States, and the World Wide Web) exhibit intriguing characteristics consistent with what they call small-world structure (Watts & Strogatz, 1998). Specifically, such networks are sparsely connected, show strong local clustering, and possess short average path-lengths between nodes. Steyvers and Tenenbaum were interested in whether semantic networks conform to small-world structural principles. Hence, they examined the degree of connectivity in three large-scale databases, *Roget’s Thesaurus of English Words and Phrases* (Roget, 1911), *The University of Florida Word Association Norms* (Nelson, McEvoy, &
Schreiber, 1998), and *WordNet* (Miller, 1990), and computed connectivity measures that reflect how densely a particular word is connected to other words. Networks were then created from these norms. For example, using the Nelson et al. norms, Steyvers and Tenenbaum connected two words in an undirected manner if one of the words was generated as a free associate to the other word. Once the networks are formed, it is possible to compute a set of metrics that capture the organization of the network, e.g., the number of nodes, the number of connections, clustering coefficients, and average path length between nodes. Using these metrics, Steyvers and Tenenbaum determined that semantic networks indeed exhibit a small-world structure - most words are sparsely interconnected, the average path length between words is short, and there is a high degree of local clustering. Interestingly, a few nodes are highly interconnected (*hubs*), and these hubs afford relatively short path lengths within the network. An example of the hubs and connections described by Steyvers and Tenenbaum is shown in Figure 2.

![Figure 2. Partial semantic network derived from free association. Each directed edge reflects an association between a cue and response.](image-url)
For our semantic connectivity analysis, we entered two connectivity measures based on Nelson et al.’s word association norms and Miller’s WordNet in Step 3 of the regression analysis. Both connectivity measures yielded reliable facilitatory effects in lexical decision ($\beta_{\text{Word Association}} = -0.21, p < 0.001; \beta_{\text{WordNet}} = -0.07, p < 0.01$). In naming, the influence of WordNet connectivity was weakly facilitatory ($\beta_{\text{WordNet}} = -0.04, p < 0.10$). It is interesting that two quite different definitions of meaning, imageability and connectivity, yielded such convergent results. The significant effects of connectivity also suggest that readers’ semantic networks may be characterized by a small-world structure. It is important to note that Steyvers and Tenenbaum also reported the predictive power of these measures in naming and lexical decision performance; however, there were fewer predictors included in their paper.

In order to assess the reliability and generalizability of the findings discussed above, we replicated our analyses with items selected from the English Lexicon Project (ELP) database (http://elexicon.wustl.edu), a repository of lexical decision and naming data for over 40,000 words and 1242 participants (Balota et al., 2002). The ELP represents a collaboration between six universities to provide the lexical and behavioral characteristics of the majority of words that are recognized by a typical college undergraduate. Despite the methodological diversity underlying the data set (heterogeneity of participant populations and testing environments), the ELP replication was remarkably consistent with the megastudy’s findings, yielding standardized regression coefficients that were comparable in magnitude and direction (see Table 1). In both datasets, semantic effects were reliably facilitatory even when confounding variables are controlled for, and were stronger in lexical decision than in naming.
It is important to remember that these effects were significant after other variables were partialed out in the first two steps of the regression analyses. The stronger semantic effects observed in lexical decision are not too surprising, since the task emphasizes meaningfulness as a dimension for carrying out word/nonword discriminations. More interestingly, one also observes reliable semantic effects in speeded naming, a task that does not require meaningful words to be discriminated from non-meaningful nonwords. Taken together, the results support the notion that meaning is activated very early in word recognition via cascadic mechanisms, and interactively contributes to the processes culminating in a lexical decision or naming response (see Figure 3). Furthermore, it is apparent that there are large task-specific effects for semantic variables, with stronger effects for lexical decision than naming. Borrowing terminology from the memory literature, this is consistent with the argument that process-pure measures (Jacoby, 1991) of lexical processing do not exist; different tasks assess different aspects or stages of the lexical system.
VI. TWO-WAY STREET AND THE FLEXIBLE LEXICAL PROCESSOR

As the preceding discussion makes clear, semantic effects are stronger in lexical decision than naming. The notion of task-appropriate processing and the flexible lexical processor is a useful heuristic for accommodating these results. Figure 3 schematically describes how task-appropriate processing and attentional control might influence different aspects of the lexical processing system. Of course, this is merely used for illustrative purposes and is not intended to be portrayed as a model of lexical processing.

![Figure 3. The flexible lexical processor.](image)

In this framework, a number of distinct processing modules/pathways subserve the computations involved in orthography, phonology, meaning, syntax, and high-level discourse integration (Balota, Paul, & Spieler, 1999), with the specific goals of the task directing attention to the appropriate processing dimensions. Of course, these different processing pathways clearly are available to people (along with many more such as the color, grammatical class, spatial location). The influence of these pathways is modulated by attentional control systems that are engaged by the experimental task demands. For
example, speeded naming, which requires spelling-to-sound conversion is primarily
driven by the connections between the orthography and phonology modules. In contrast,
in lexical decision, participants discriminate between meaningful words and relatively
meaningless nonwords (Balota & Chumbley, 1984), and hence are more likely to
emphasize the computations between the orthography and meaning. This is consistent
with the larger semantic effects in lexical decision than in naming.

Of course, bringing on-line different processing components of performance in
anticipation of task demands is a two-way street. One should not only obtain stronger
semantic effects in lexical decision; one would also expect stronger phonological (e.g.,
spelling-to-sound consistency) effects in naming. Figure 4 reports the mean regression
coefficients from the second and third steps of the large-scale regression analyses from
both the megastudy and the same items from the English Lexicon Project presented
above.
There are two things to note about these results. First, there is considerable consistency across the two datasets. Hence, if one is concerned about the reliability of the results from a given large-scale dataset, it appears that these concerns are unwarranted. The current pattern is particularly intriguing in that the items in the megastudy were only single syllable words, whereas the items from the English Lexicon Project were the same single syllable words embedded among multisyllabic words, proper names and even contractions. Second, there is relatively little consistency in the size of the regression coefficients across lexical decision and naming performance. Specifically, lexical decision is driven primarily by familiarity and meaning, whereas, naming is primarily
driven by length, orthographic neighborhood size, and spelling/sound consistency measures. At this level, it appears that participants are bringing on-line a very different set of processes depending upon the task demands.

*How much flexibility is there?*

An important question that one must ask is how much control there actually is in the system. For example, can one produce more subtle effects of attentional control? Control ultimately depends on the strength of the pathway that needs to be controlled. For example, researchers originally argued that the priming effects from short SOA semantic priming studies (e.g., Neely, 1977) and the interference of the word dimension in Stroop performance were outside the influence of control within a typical experimental setting, because these pathways were so well-developed. Such findings were used to support the automatic/modular aspect of the lexical processing system. However, subsequent research has indicated that even short SOA priming effects (e.g., Balota, Black & Cheney, 1992) and the interference effects in Stroop (e.g., Besner, 2001) are sensitive to control manipulations. Interestingly, there is even evidence that masked repetition priming effects can be modulated by context manipulations (see Bodner & Masson, 2004). Hence, the classic evidence for uncontrollable pathways, i.e., modular automatic systems, has clearly been questioned.

One area where there has been considerable interest in control systems in word recognition is the extent to which subjects can exert control over lexical and sublexical pathways in naming performance. For example, in one study by Balota, Law, and Zevin (2000), the *regularization* condition required participants to name a set of words and nonwords according to the spelling to sound principles in the language (i.e., pronounce
PINT such that it rhymes with HINT). In a second condition, participants were given normal naming instructions. As one might suspect, participants had some difficulty naming exception words under the regularization instructions (mean accuracy was .65 and correct response latencies were on the order of 1 second). More interestingly, however, was performance on a set of regular words, which were correctly named in an identical fashion in both the regularization and normal naming conditions. These items produced a typical word-frequency effect (23 ms) in the normal naming conditions, but actually produced a reversal of the word-frequency effect (-43 ms) in the regularization conditions. Moreover, there was no hint of a lexicality effect (words faster than nonwords) in the regularization condition (1 ms), but there was a large effect of lexical status (76 ms) in the normal naming conditions. These results suggest that one can access a sublexical spelling-to-sound route when directly instructed, albeit at a cost to accuracy and response times. One cannot totally control the lexical influence since there was a reverse word frequency effect. This could be viewed as consistent with a Stroop type effect, wherein, one finds greater interference from the prepotent high-frequency words, compared to the low-frequency words. Clearly, further research would be useful for addressing the following issues. First, can one modulate the size of the frequency effect by increasing practice on the sublexical route? Second, can meaning-level variables be modulated by such regularization instructions?

Let us now turn to a more subtle level of control. Consider for example, the possibility that participants may exert control over the relative contributions of a pathway that computes spelling-to-sound correspondences and a pathway that computes lexical-level information in a normal speeded naming task. Indeed, preliminary evidence for
strategic control in word recognition was provided by a speeded naming study that manipulated list composition (Monsell, Patterson, Graham, Hughes, & Milroy, 1992). Monsell et al. examined naming performance for high frequency exception words (words that violate English spelling-sound rules, e.g., PINT) when embedded within a list of other exception words, compared to when these items were embedded in a list of nonwords (e.g., FLIRP). When only exception words were presented, naming latencies were shorter and regularization errors (i.e., pronouncing PINT so that it rhymed with HINT) were less likely, compared to when only nonwords were presented. This finding suggests that participants were able to strategically inhibit or ignore sublexical orthography-to-phonology computations when they were anticipating only exception words, since sublexical processes will, by definition, generate incorrect pronunciations for exception words.

This issue has been further investigated by Zevin and Balota (2000), using a variation of a task developed by Midgeley-West (1979). Midgeley-West had participants read aloud a single list of 24 monosyllabic nonwords followed by an exception word (WOLF). Seven of 18 participants regularized WOLF (so that it rhymed with GOLF), suggesting that lexical knowledge may have been suppressed by the predominantly nonword context. In Zevin and Balota’s modified priming procedure, each trial consisted of six items that participants had to name. The first five items were primes (either low frequency exception words like BISCUIT or nonwords like FLIRP) and the sixth, final word was the target, which could vary on some targeted dimension. The basic prediction was that low frequency exception primes should emphasize lexical processing, while nonword primes should emphasize sublexical processing. In fact, Zevin and Balota
observed pathway influences on *lexicality, frequency, and regularity* effects. When the lexical pathway was primed, lexicality effects (words named faster than nonwords) and frequency effects (high frequency words named faster than low frequency words) were larger than when the sublexical pathway was primed. In contrast, priming the sublexical pathway produced larger regularity effects (regular words, which conform to spelling-sound rules, named faster than exception words). Most intriguingly for the present chapter, Balota and Zevin reported that the influence of imageability, a meaning-level variable, could be selectively modulated by pathway priming. Specifically, in Experiment 4, imageability effects (high imageability words named faster than low imageability words) were reliable only when the lexical pathway was primed, suggesting that the meaning/semantic module (see Figure 3) is recruited when difficult, low frequency exception words have to be named.

To establish the reliability of the imageability effect for the set of items in the Zevin and Balota study, we used the naming response latencies from the same set of items taken from the ELP behavioral database. The ELP corpus has a much larger representation of lower frequency, multisyllabic, and multimorphemic words, and one might predict that such a context will encourage a relatively heavy reliance on lexical/semantic information. Interestingly, when we examined the ELP naming latencies for the stimuli used in Zevin and Balota, we found a large imageability effect of 33 ms ($p < .01$), a finding that is consistent with our prediction.

It is important to point out that there are competing explanations for the phenomena observed by Monsell et al. (1992) and Zevin and Balota (2000). Instead of viewing list composition/context effects as evidence for strategic pathway control, some
researchers have argued that these results are more compatible with a *flexible time-criterion* hypothesis (Lupker, Brown, & Colombo, 1997; Kinoshita & Lupker, 2003). In this account, when readers are presented with a word, they do not initiate articulation as soon as they are able to. Rather, a flexible time criterion is adopted so that articulation could begin either *before* or *after* the articulatory program is fully compiled (Kinoshita & Lupker, 2003). The time criterion for articulation onset is mainly determined by the difficulty of items in a trial block; “difficult” blocks consist mainly of slow (i.e., long response latencies) items, while “easy” blocks consist mainly of fast (i.e., short response latencies) items. This implies that the same items will be responded to faster when embedded in an “easy” block than when embedded in a “difficult” block. Kinoshita and Lupker used a different context manipulation than Zevin and Balota, and investigated the same variables that Zevin and Balota investigated in their first three experiments. Kinoshita and Lupker found that target processing was influenced only by the relative difficulty of contextual primes, not by the different routes presumably engaged by the primes *per se*. Kinoshita and Lupker did not address the finding that imageability effects can be modulated by pathway manipulations. It should also be noted that Kinoshita, Lupker, and Rastle (2004) have recently provided evidence that one can indeed modulate the lexicality (but not the regularity) effect via list context manipulations. In addition, Reynolds and Besner (2005) have recently demonstrated using a switching task that one can find lexical and sublexical pathway switching above and beyond any response latency criterion effects. Hence, there does appear to be converging evidence from other paradigms of some level of pathway control.
If Kinoshita et al. were correct, and the Balota and Zevin results were due to the difficulty of the items in the priming conditions, as reflected by the latencies for the priming items, then there should be an easy way to statistically control for this possibility in this dataset. That is, by covarying out the response latencies on the prime trials, one can determine if there is still an effect of condition. In fact, we have recently conducted this analysis, and route priming modulated lexicality, regularity, and imageability effects even after prime response times were controlled for. Hence, it appears that setting one’s criterion based on the speed of context items is an insufficient explanation of the Zevin and Balota results.

The current status of the route priming results suggests that it is difficult to modulate the lexical and sublexical contributions to naming performance via simple context manipulations. As Zevin and Balota (2000) argued, such effects are likely to depend on the strength of the manipulations. In fact, Zevin and Balota developed the within-trial priming procedure to increase the strength of the manipulation. Although the debate between pathway control and time-criterion is clearly important, it is perhaps misleading to view the two hypotheses as mutually exclusive. We believe that it is likely that both mechanisms work in tandem, and as pointed out earlier, one way to empirically test this is to determine if pathway selection effects are reliable after the difficulty of the list context has been controlled for.

To summarize, it appears that meaning-level influences are not invariant in speeded naming performance. Instead, imageability effects can be modulated as a function of list context (Zevin & Balota, 2000). However, it is also likely that one needs a strong manipulation that encourages lexical or sublexical pathways before such effects
can be observed. We want to reiterate the utility of a flexible lexical processor framework, where attentional control can selectively influence the output from different processing pathways. Importantly, while contextual factors may influence the operation of the attentional control mechanism, we acknowledge that the degree of attentional control is likely to depend on factors such as: (1) the ability of contextual factors to engage an appropriate task set, (2) the ability to maintain a representation of this task set over time, and (3) the prepotent strength of the pathways.

VII. MEANING VS. CONTROL: EVIDENCE FROM DEMENTIA OF THE ALZHEIMER’S TYPE AND SEMANTIC DEMENTIA

Examining different types of language breakdowns across distinct clinical populations may afford another converging line of evidence concerning how control mechanisms may interact with structural systems. For example, there is currently some debate regarding the nature of the language breakdowns that occur in early stage Dementia of the Alzheimer’s Type (DAT). Individuals with early stage DAT produce large deficits in tasks such as picture naming and category fluency. Some researchers (e.g., Salmon, Butters, & Chan, 1999) have suggested that this reflects a breakdown in the integrity of the semantic network, whereas others (e.g., Balota, Watson, Duchek, & Ferraro, 1999; Nebes, 1989; Ober, 1995) have argued from intact semantic priming effects that such breakdowns are more likely due to changes in the control systems used to access the semantic network. The distinction between these two positions is not only theoretically important but has applied implications when considering the types of tasks one wishes to employ to identify individuals with early stage DAT.
One way to adjudicate between these accounts is to compare individuals who have a clear semantic deficit against individuals who have DAT. Recently, we have been exploring a group of individuals who have a variant of fronto-temporal lobe dementia referred to as Semantic Dementia (see Hodges, Garrard, & Patterson, 1998, for a review of this literature). Figure 5 (reproduced from Cortese, Balota, Sergent-Marshall, & Buckner, 2003) displays performance on a group of healthy control, DAT, and Semantic Dementia (SD) individuals on a set of semantic and non-semantic tasks. As one can see, in general, performance for most participants on semantic and non-semantic tasks is positively correlated. However, this figure also shows that there are a few participants who showed disproportionate breakdowns in semantic tasks. We have since classified 3 of the 4 participants as having SD.

Figure 5. Scatter plot of semantic and non-semantic z-scores of 83 participants who participated in the Cortese et al. (2003) study. The line represents the best-fit to the data with the four PSI participants excluded ($r = 0.58$).
Gold, Balota, Cortese, Sergent-Marshall, Snyder et al. (2005) explored speeded naming performance in individuals with SD and those who have DAT. Gold et al. argued that if there was a deficit in the lexical/semantic route then one should find disproportionate length effects in these individuals. Such effects would be indicative of a greater reliance on a more serial spelling-to-sound computation. Indeed, Weekes (1997) has provided evidence that nonwords (items that are semantically barren) produce large length effects, whereas words produce very little if any length effect when other extraneous variables are controlled. Hence, one might expect individuals with SD to treat words as nonwords, and hence produce large length effects. Likewise, if individuals with DAT also have such a breakdown in the semantic/lexical route, they should at least produce larger length effects compared to healthy age-matched control individuals. The results from the speeded naming task are displayed in Figure 6.

![Figure 6. Mean group naming latencies as a function of letter length.](image-url)
As shown here, individuals with SD produce exaggerated length effects, compared to both the healthy control individuals and the individuals with early stage DAT, with no difference between the latter two groups. Because the DAT individuals produced similar length effects as the healthy controls, and individuals with a clear semantic deficit produce exaggerated effects of length, one can argue that the breakdown is not in the lexical/semantic structure, at least in early stage DAT.

A similar conclusion was made by Cortese et al. (2003) from a homophone spelling task. Cortese et al. compared spelling performance on a set of items in which both the sound-to-spelling dominance and the meaning dominance of the auditorily presented word converged on the same spelling (e.g., /weist/ should be spelled WASTE both based on sound-to-spelling rules and meaning dominance). Alternatively, there was a set of items in which the sound to spelling dominance diverged from the meaning dominance (e.g., /plein/ should be spelled PLAIN according to sound to spelling dominance and PLANE according to meaning dominance). Although there was little difference in spelling the items that converged on meaning and spelling dominance across the participants, there was a clear difference in performance on the items that diverged on meaning and spelling dominance. In particular, as shown in Figure 7, the DAT individuals and healthy controls were equally likely to rely on meaning dominance in spelling the homophones (i.e., spelling /plein/ like PLANE), whereas the individuals with SD were much more likely to rely on the spelling dominant pathway (i.e., spelling /plein/ as PLAIN).
Cortese et al. also compared spelling performance on a set of words that had consistent sound-to-spelling correspondence (e.g., SLICK), and a set of words that had inconsistent sound-to-spelling correspondence (e.g., PLAID). For inconsistent words, there is a conflict between the correct spelling (lexical pathway) and the regularized spelling (sublexical pathway), and the system needs to exert control, i.e., suppress the sublexical sound-to-spelling route to output a correct spelling. Interestingly, as shown in Figure 8, the individuals with DAT produced a reliable impairment in spelling the inconsistent words, compared to the healthy age-matched controls.
It is noteworthy that a similar pattern is found in speeded naming performance wherein DAT individuals are more likely to regularize irregular words compared to healthy age-matched control individuals (see Balota & Ferraro, 1993; Patterson, Graham, & Hodges, 1994). Instead of considering this pattern as a reflection of semantic memory degradation, we believe that this may at least in part be due to a breakdown in controlling the output from a sublexical pathway, which needs to be suppressed in this context.

The results reviewed in this section indicate that it is useful to again consider the task constraints when making inferences about the locus of observed semantic deficits. In particular, the breakdowns that DAT individuals exhibit in supposed semantic tasks such as picture naming and category fluency could be signaling deterioration in task-specific control systems rather than the integrity of the semantic network. Comparing groups of individuals who have clear semantic memory loss (individuals with SD) against healthy...
older adults and individuals with early stage DAT provides leverage in understanding the locus of such deficits.

**VIII. TASK SET AND NEUROIMAGING**

The present chapter has emphasized the argument that one needs to consider task-appropriate processing to understand the manner in which a variable influences performance in a lexical task. We end with a few observations from the neuroimaging domain regarding attempts to provide leverage on such task signals. Of course, the field of cognitive neuroimaging has developed considerably since the seminal work of Petersen, Fox, Posner, Mintun and Raichle (1989), who were the first to observe remarkably distinct regions of the brain activated as a function of task set. In particular, positron emission tomography (PET) scans of participants while processing single words indicated that, compared to baseline fixation performance, occipital areas were engaged for passive viewing of words, temporal areas were engaged for reading words aloud, and frontal regions were involved when participants generated verbs to nouns. Since this classic work, there have been many further developments. For example, researchers began tracking manipulations within the same task. In this vein, Fiez, Balota, Raichle, and Petersen (1999) used PET imaging to examine responses as participants read lists that varied as a function of frequency, consistency, and lexicality. In this study, there was clear evidence of a region in the left medial inferior frontal region (near BA 45) that was particularly sensitive to low-frequency inconsistent words.

More recently, Palmer, Brown, Petersen, and Schlaggar (2004) have provided an updated review of imaging work pertaining to the functional neuroanatomy of single word reading. Activations in different brain regions are reliably associated with different
language functions. For example, orthographic processing is associated with the temporal lobe its junction with the temporal lobe (Petersen et al., 1989), spelling-to-sound conversion is associated with the left inferior frontal region (Fiez et al., 1999), phonological decomposition is associated with the left posterior temporal lobe (Simos, Breier, Fletcher, Foorman, Castillo et al., 2002), and semantic processing is associated with both the temporal lobe (between superior and middle temporal gyrus) and the inferior frontal regions (Petersen et al., 1989; Price, Moore, Humphreys, & Wise, 1997).

One important limitation that applies to both PET studies and early functional magnetic resonance imaging (fMRI) studies is that blocked designs were mandatory. As a result, neural activity may reflect either item-specific processing, or task-related strategic states that are sustained across the trials of a task; item and state effects are completely confounded in blocked designs. Of course, one of the main points of the current chapter is that it is critical to decompose these two components of performance. This problem has been partly remedied with the advent of event-related fMRI designs (Dale & Buckner, 1997). Using event-related fMRI, one can now extract the fMRI BOLD (blood oxygen level dependent) response for specific trials. Although this was an important advance, event-related fMRI is still insensitive to task-related effects that may be sustained across trials. Fortunately, there has been progress in developing an fMRI design that is a hybrid of block and event-related designs. In the mixed block/event-related design (Donaldson, 2004), one is able to separately extract brain signals that are sustained across task trials (state effects), and transient brain signals that are associated with specific trial events (item effects). State effects, which are potentially related to task-level control signals
(Visscher, Miezin, Kelly, Buckner, Donaldson, et al., 2003) may afford important leverage in understanding task-related attentional modulations.

Although mixed designs are relatively new, there is compelling evidence for their efficacy. For example, Otten, Henson, and Rugg (2002) conducted a mixed fMRI study where participants made semantic or phonological decisions about visually presented words. After statistically controlling for item-related activity, they found that the semantic encoding condition was associated with sustained activity in the inferior medial parietal and left prefrontal cortex, whereas the phonological encoding condition was associated with sustained activity in the medial parietal cortex. Interestingly, the magnitude of sustained brain activity predicted subsequent memory performance. These results are consistent with the notion that the different encoding instructions induced different ‘encoding modes’ in participants; these modes are evinced by sustained activity in brain regions that support memory encoding (Donaldson, 2004). Other studies have successfully used the mixed design to better understand the neuroanatomical bases of cognition in domains as diverse as memory encoding (Reynolds, Donaldson, Wagner, & Braver, 2004), memory retrieval (Velanova, Jacoby, Wheeler, McAvoy, Petersen et al., 2003), and task switching (Braver, Reynolds, & Donaldson, 2003).

There has not yet been much systematic work examining state and item effects in visual word recognition (but see Palmer, 2003). However, the mixed design is potentially a powerful tool for elucidating, at the neuroanatomical level, the task-related effects we have been discussing in this chapter. For example, the task-appropriate processing framework predicts that one might identify item effects that generalize across different reading tasks but are modulated by the relevance to the task demands, and state effects
that are specific to the reading task being employed. In this way, one may be able to better understand how the brain flexibly implements operations to accomplish a given task goals.

Conclusion

The present chapter has had three interrelated goals. First, at a theoretical level, we have attempted to elucidate the consequences of particular stances taken regarding the influence of meaning on lexical processing tasks. In this light, we have provided evidence that meaning-level variables do indeed play a role in both lexical decision and word naming, with smaller effects in the later task. Hence, we question the utility that there is a measurable point in time where a lexical representation has been accessed, but meaning has not become available. Second, we have reviewed a series of methodological considerations that may afford some progress in understanding how processes evolve across time to influence response time distribution and the actual implementation of the response. Along these same methodological considerations, we have attempted to show how large-scale databases can provide converging evidence concerning the influence of lexical variables on word recognition. Finally, we have emphasized the importance of task-specific operations and a flexible lexical processor, and have discussed the potential utility of exploring differences in process and structure across different populations and via recent neuroimaging techniques. We believe that ultimately, any model of word recognition will need to incorporate both structural mechanisms and attentional mechanisms that bring to fore a configuration that is tuned to accomplish the goals of a given word recognition task.
References


Cutler, A. (1981). Making up materials is a confounded nuisance, or: Will we be able to run any psycholinguistic experiments at all in 1990? *Cognition, 10*, 65-70.


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


