Psychological Review
2006, Vol. 113, No. 1, 181–195

0033-295X/06/$12.00 DOI: 10.1037/0033-295X.113.1.181

Parallel Distributed Processing and Lexical–Semantic Effects in Visual Word Recognition: Are a Few Stages Necessary?

Ron Borowsky
University of Saskatchewan

Derek Besner
University of Waterloo

D. C. Plaut and J. R. Booth (2000) presented a parallel distributed processing model that purports to simulate human lexical decision performance. This model (and D. C. Plaut, 1995) offers a single mechanism account of the pattern of factor effects on reaction time (RT) between semantic priming, word frequency, and stimulus quality without requiring a stages-of-processing account of additive effects. Three problems are discussed. First, no evidence is provided that this model can discriminate between words and nonwords with the same orthographic structure and still produce the pattern of factor effects on RT it currently claims to produce. Second, the level of representation used by the model to make a lexical decision is inconsistent with what is known about how skilled readers with damage to their semantic system make word/nonword discriminations. Finally, there are a number of results that are difficult to reconcile with the single mechanism account. The authors' preference is to retain the stages-of-processing account.

Keywords: stages-of-processing, parallel distributed processing (PDP), lexical–semantic priming, word frequency, stimulus quality

Overview

The idea that one can make inferences about the organization and nature of underlying mental processes by examining the pattern of factor effects on the time to respond in various tasks has a long and distinguished history in cognitive psychology. In particular, the observation that two or more factors have additive effects on reaction time (RT) has traditionally been interpreted as evidence that these different factors affect separate, serially organized stages of processing (Sternberg, 1969, 1998; see also McClelland, 1979 for how cascaded processing can also yield additive effects of two factors on RT and Roberts & Sternberg, 1993 for limitations to this claim). The present commentary considers an account proposed by Plaut and Booth (2000). They argue that a single mechanism embedded in an implemented parallel distributed processing (PDP) model with a sigmoid activation function correctly simulates the results of multifactor RT experiments in which the effects of some factors interact (e.g., A and B; B and C) whereas other factors have additive effects (e.g., A and C), without the need to attribute the effects of the additive factors (A and C) to separate stages. The analysis of Plaut and Booth's computational model provided here suggests that this single mechanism account fails in several ways. We therefore continue to prefer the traditional separable stages-of-processing view of how additive effects of factors on RT are best understood.

The Plaut and Booth (2000) model is also potentially important for another reason. In earlier work, Seidenberg and McClelland (1989) claimed that their implemented PDP model, which was trained using all the monosyllabic words in English, produced lexical decision accuracy on a par with that seen by skilled human participants. Unfortunately, closer examination of the Seidenberg and McClelland model showed that its lexical decision accuracy was, in a particular context, actually quite poor (Besner, Twilley, McCann, & Seergobin, 1990). A number of PDP models of visual word recognition have been subsequently published, and it is clear that lexical decision is not a trivial process to simulate (e.g., Borowsky & Masson, 1996; Joordens & Besner, 1994; Plaut, 1997). Plaut and Booth's (2000) model is therefore potentially important because it could be viewed as an existence proof for the claim that a PDP model can produce accurate lexical decision performance while also simulating effects associated with timed lexical decision performance. In contrast, we argue that the model was not properly assessed and therefore no evidence exists that this model can produce accurate lexical decision performance while also simulating the joint effects of various factors on timed lexical decision performance.

It is also argued that the level of representation that the model uses to make a lexical decision does not correspond to the level that humans use at least some of the time for this task (as Plaut & Booth, 2000, themselves concede, see p. 812). Our general conclusion is that, in its currently implemented form, this model does little in the way of advancing our understanding of how humans do

Ron Borowsky, Psychology Department, University of Saskatchewan, Saskatoon, Saskatchewan, Canada; Derek Besner, Psychology Department, University of Waterloo, Waterloo, Ontario, Canada.

This research was supported by grants from the Natural Sciences and Engineering Research Council of Canada to Ron Borowsky and Derek Besner. We thank Max Coltheart, Mike Masson, and Dave Plaut for constructive comments.

Correspondence concerning this article should be addressed to Ron Borowsky, Psychology Department, University of Saskatchewan, 9 Campus Drive, Saskatoon, Saskatchewan, S7N 5A5 Canada, or Derek Besner, Psychology Department, University of Waterloo, Waterloo, Ontario, N2L 3G1 Canada. E-mail: ron.borowsky@usask.ca or dbesner@watarts.uwaterloo.ca

what they do so easily: discriminate words from nonwords with the same orthographic structure.

Introduction

Skilled readers have very little metacognitive awareness of the underlying perceptual and cognitive processes involved in recognizing a printed word's orthographic form and retrieving its meaning. However, the efficiency and speed with which we are capable of reading for meaning has sometimes been taken to imply that a single mechanism (i.e., a single process) is sufficient to account for processing at the single word level. Indeed, an early theory of word identification subscribed to the idea of a singular "mental lexicon," in which the representations of orthography (spelling), phonology (pronunciation), and semantics (meaning) were all packaged together for words known to the reader (e.g., Treisman, 1960). However, over the last 4 decades, a number of empirical observations in both normal participants and those with acquired brain damage have led a wide range of investigators to suppose that multiple stages of processing are needed to account for lexical–semantic processing effects (e.g., Becker, 1979; Becker & Killion, 1977; Besner & Smith, 1992; Borowsky & Besner, 1993; Brown & Besner, 2002; Coltheart, Curtis, Atkins, & Haller, 1993; Coltheart, Rastle, Perry, Langdon, & Zeigler, 2001; Forster & Davis, 1984; Grainger & Jacobs, 1996; McClelland & Rumelhart, 1981; McClelland, 1987; Morton, 1969, 1979; Morton & Patterson, 1980; Neely, 1977, 1991; Stanners, Jastrembski, & Westbrook, 1975; Stolz & Besner, 1996, 1998; see also Carr & Pollatsek's, 1985, review).

In contrast, Plaut and Booth (2000) have argued that, in the context of the lexical decision task, a single mechanism embedded in a parallel distributed processing model is sufficient to account for both semantic priming effects and lexical processing effects. We begin by outlining some of the empirical findings that the Plaut and Booth model purports to explain, and we follow this with a description of their model.

The Phenomena

There are many phenomena that any successful account of lexical–semantic processing in the context of the lexical decision task needs to be able to explain. We restrict our attention here to only a few of these phenomena because they suffice to illustrate what Plaut and Booth (2000) consider to be some of the strengths of their approach. We begin by describing the phenomena that have been reported in the lexical decision literature that will be considered here in the context of the Plaut and Booth model.

The Joint Effects of Semantic Priming and Stimulus Quality

One well-documented phenomenon is the *semantic priming effect* in which a target word (e.g., DOG) is identified more quickly when preceded by a related prime (e.g., CAT) compared with an unrelated prime (see Neely's, 1991, review). This semantic priming effect sometimes interacts with the effect of stimulus quality. That is, a reduction in stimulus quality slows the processing of unrelated targets more than it slows the processing of related targets (e.g., Becker & Killion, 1977; Borowsky & Besner, 1991, 1993; Meyer, Schvaneveldt, & Ruddy, 1975). There are also conditions under which these same two factors produce additive effects on lexical decision RT that are within the same range of RTs that showed an interaction in some of the former studies (Borowsky & Besner, 1991, 1993; Brown & Besner, 2002; Stolz & Neely, 1995).

The Joint Effects of Stimulus Quality and Word Frequency

The second phenomenon involves the *word frequency effect* in which high-frequency words are identified more quickly than low-frequency words (e.g., Forster & Chambers, 1973; see also Monsell's, 1991, review). Stimulus quality does not differentially affect the processing of high- and low-frequency words. Instead, a reduction in stimulus quality produces only a main effect on RT (e.g., Balota & Abrams, 1995; Becker & Killion, 1977; Borowsky & Besner, 1993; Plourde & Besner, 1997; Stanners et al., 1975).

The Joint Effects of Context and Word Frequency

The third phenomenon is that context interacts with word frequency. The semantic priming effect is larger for low-compared with high-frequency words (Becker, 1979; Borowsky & Besner, 1993).

The Plaut and Booth (2000) PDP Model

A written word is represented by a pattern of activation over a set of 18 orthographic units that encode three-letter inputs (words are composed of consonant–vowel–consonant [CVC] strings), and its meaning is represented by another pattern of activation over a set of 100 semantic units. A set of 100 hidden units receives input from the orthographic units and is fully interconnected to the semantic units, which are fully connected to each other (see Figure 1).

Context is implemented by varying both co-occurrence probabilities (i.e., associative relatedness, or how often one word tends to occur with another word) and semantic feature overlap (i.e., semantic relatedness, or similarity of features between exemplars of a category). Finally, word frequency was simulated by varying the frequency with which words were presented during training.

Plaut and Booth's (2000) model is based on an earlier model (Plaut, 1995), which was trained to map written words to their

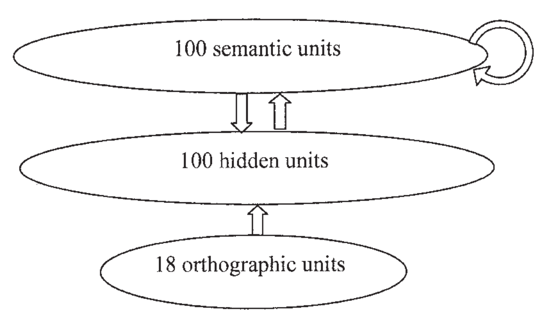

Figure 1. The architecture of the Plaut and Booth (2000) model.

meanings. The Plaut (1995) model had been claimed to exhibit the pattern of joint effects between stimulus quality, context, and word frequency reported by Borowsky and Besner (1993). The Plaut (1995) model is clearly the foundation for the Plaut and Booth (2000) model, and we note that it is referred to often throughout the discussion of the 2000 model. For example:

> "Our account takes as its starting point a preliminary distributed network simulation developed by Plaut (1995)" (Plaut & Booth, 2000, p. 789). "The approach taken is closely related to the one used in the Plaut (1995) simulation" (Plaut & Booth, 2000, p. 800). "The semantic representations of words were the same as those used by Plaut (1995)" (Plaut & Booth, 2000, p. 801).

In Plaut (1995), the description of how stimulus quality is implemented is also clearly stated:

> "Empirical studies have found that priming is increased if targets are degraded visually (e.g., by reducing contrast; see Neely, 1991). To investigate this effect in the model, the orthographic input patterns for targets were reduced in visual contrast by scaling the input values toward the neutral value. . ." (Plaut, 1995, p. 41).

Both of these variants of the model manipulate the strength of the external orthographic input; Plaut and Booth (2000) discussed it as a means of modeling perceptual ability of the participants, and Plaut (1995) discussed it as a means of modeling perceptual quality of the stimulus. Thus, without claiming that it is an exhaustive measure, we wish to draw the reader's attention to the fact that orthographic input strength serves as an implementation of both perceptual ability of the participant, and perceptual quality of the stimulus.

Three Problems for the Current Implementation of the Model

Problem one: Does the model discriminate between words and well-formed nonwords? The lexical decision task is typically conceived of as one that forces participants to interrogate their mental lexicon (e.g., is "mantiness" a word? See Norman, 1969, p. 162). The vast majority of investigators have therefore used nonwords that preserve the orthographic structure that is seen in the word stimuli. If the structure of the nonwords differs too much from the structure of the words, then participants may, at least some of the time, develop and use strategies that are restricted to the orthographic and/or phonological level (e.g., see Borowsky & Masson, 1996; Shulman, Hornak, & Sanders, 1978). Unless one is trying to manipulate the level at which participants are monitoring activation to make their lexical decisions, using orthographically strange nonwords is counterproductive because it defeats the intent of the task, which is to probe lexical/semantic organization and processing dynamics. If a model is being evaluated with respect to its ability to simulate "lexical decision" as typically conceived, then the stimuli that it is tested with have to be faithful to the stimulus characteristics that are generally used in experiments on skilled readers.

A central problem here thus concerns how Plaut and Booth (2000) tested the model's lexical decision ability. Plaut and Booth's simulation of lexical decision used CVC letter strings constructed from 10 consonants and 5 vowels; there were 500 possible strings. Plaut and Booth chose a random 128 words and trained their orthography-to-semantics model on these items. Once the model was fully trained, an obvious way of testing its lexical decision performance would have been to test whether the model could classify the 500 possible letter strings into "words" (the 128 letter strings it was trained on) and "nonwords" (the 372 letter strings it had never seen before). But Plaut and Booth did not do this. Instead, they used 128 letter strings with a vowel–consonant–vowel (VCV) orthographic structure, a structure to which the model had never been exposed. There is thus no evidence in the Plaut and Booth (2000) paper that this model can discriminate between words and nonwords at a high level of accuracy when the nonwords have the same orthographic structure as the words.

It might of course be that the model could do this if tested appropriately, we simply do not know. Simulation accounts are meant to provide an existence proof. Given that there are a number of demonstrations that some computational models behave in ways unanticipated by the modelers (e.g., Besner & Roberts, 2003; Besner et al., 1990; Reynolds & Besner, 2002, 2004), it is important to run the simulations with stimuli that are typically used in experiments with human readers.

Of course, PDP models have been challenged previously concerning their ability to make accurate lexical decisions. Seidenberg and McClelland (1989) were the first to claim that their PDP model was capable of high lexical decision accuracy, but closer examination by Besner and colleagues (1990) showed that this was not the case. It is thus important to acknowledge that Plaut (1997) was able to show that a PDP model can discriminate very well between words and nonwords with the same orthographic structure based on a measure of familiarity defined over semantics. However, there are some differences between the Plaut (1997) model and its performance and the present model.

The Plaut (1997) model differs from the Plaut and Booth (2000) model in that the former has twice as many semantic units (200 vs. 100), 10 times the number of hidden units (1,000 vs. 100), and six times the number of orthographic units (108 vs. 18) than the latter. Further, the Plaut (1997) model is feed-forward only, whereas the Plaut and Booth (2000) model is recurrent. It might be that the Plaut and Booth (2000) model would perform at equivalent levels of accuracy as the Plaut (1997) model when the untrained words from the corpus are considered to be nonwords, as suggested above. This is an empirical question that remains to be answered. Further, it remains to be seen whether a high level of accuracy with this set of "nonwords" would still allow the Plaut and Booth (2000) model to produce the same pattern of effects on RT that it does now.

We do not wish to ignore the fact that Plaut and Booth (2000) offered a rationale for why they chose to test this model with words and nonwords that differed so profoundly in orthographic structure. Their argument was that the experimental data from adults and children that they were trying to simulate involved words and nonwords that were not orthographically matched because they differed in terms of bigram and trigram frequencies. They went on to suggest that the relative ratio of word to nonword similarity (in the experiments vs. the simulation) is actually quite similar (1.31 for the experiments vs. 1.19 for the simulation). This rationale is unconvincing because it ignores two larger issues. First, Plaut and Booth argued that the model simulates more than just the experiments reported in their paper. Instead, the model is repeatedly argued to provide a general account of a rather large number of results from experiments in the visual word recognition literature on skilled readers. Relatedly, such experiments with humans typically do not make the word and the nonwords as structurally different as used in these simulations for a very simple reason. Faced with words that always begin with a consonant and nonwords that always begin with a vowel, lexical decision by skilled readers could be 100% accurate by simply classifying the first letter as a vowel or a consonant.

Problem two: At what level are lexical decisions made in the model and by humans? The Plaut and Booth (2000) model makes lexical decisions at the semantic level. A serious difficulty here is that using the semantic level to make such a decision may render this form of the model incapable of explaining how literate participants with acquired, and severe damage to their semantic system can make lexical decisions with accuracy that is as good as controls without such damage. Coltheart (2004) has reviewed a number of such cases in considerable detail. He argues that these patients make lexical decisions based on the activation of a lexical entry in a system in which words are represented by individuated nodes that are distinct from semantic representations. This parallels the view in some models of how normal participants can make lexical decisions (e.g., Coltheart et al., 2001).[1] When there is enough activation to meet some criterion, the participant responds "yes." The participant responds "no" on the basis of a flexible deadline. If activation has not reached the criterion for a "yes" response by some time (*t*), then a "no" response is made (e.g., Coltheart, Davelaar, Jonasson, & Besner, 1977; Coltheart et al., 2001).

One response that could be made to this point is that patients with a damaged semantic system might produce excellent lexical decision performance not because they have consulted an orthographic lexicon, but because they performed lexical decision by reference to letter string typicality (Rogers, Lambon Ralph, Hodges, & Patterson, 2004). The claim here is that words and nonwords might not be matched for typicality when they are operationalized as bigram and trigram frequency and that participants might pick up on this fact as a basis for making a lexical decision. However, Coltheart (2004) discussed five patients with severe semantic damage whose lexical decision performance was normal despite the fact that none of these patients' performance could be explained in terms of using stimulus typicality instead of lexicality (Patient DC: Lambon Ralph, Ellis, & Franklin, 1995; Patient LR: Lambon Ralph, Sage, & Ellis, 1996; Patient SA: Ward, Stott, & Parkin, 2000; Patient MG: Rogers et al., 2004; and Patient LS: Rogers et al., 2004).

Another response is to suppose that performance by brain-damaged patients is of little or no interest to the PDP enterprise. We are confident that Plaut and Booth (2000) do not subscribe to this view, given the elaborate attempt by Plaut, McClelland, Seidenberg, and Patterson (1996) to simulate word and nonword naming performance by patients with acquired surface dyslexia and Plaut and Shallice's (1993) extensive paper on the simulation of acquired deep dyslexia.

We are therefore inclined to accept Coltheart's (2004) conclusion that any PDP model that can only make word/nonword decisions by reference to the semantic level does not provide an account of lexical decision by patients with severely impaired semantics. To be sure, we are not claiming that it is impossible for a PDP model to accomplish this. For example, one reviewer

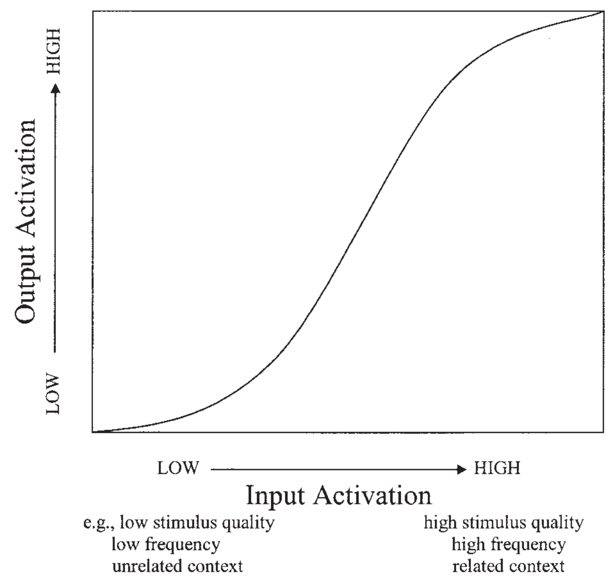

Figure 2. The sigmoid activation function from the Plaut and Booth (2000, see also Plaut, 1995) model. Output from the semantic system (representing activation, assumed by Plaut and Booth to map linearly onto reaction time) is a sigmoid function of the strength of input to the semantic system (i.e., output from the orthographic system).

suggested that orthographic familiarity sufficient to discriminate between words and well-formed nonwords could be captured by the orthographic units of the Plaut and Booth (2000) model if those units were to be set up to allow learning. Nonetheless, we admit to a certain skepticism here, given that subsequent to the failure of Seidenberg and McClelland's (1989) model to produce accurate lexical decisions in the absence of a semantic system, no implemented PDP model has been published to date that accurately discriminates between words and well-formed nonwords either in the absence of a semantic system or in the presence of a substantially damaged semantic system (as damage to this system is surely a matter of degree). If it is so easy to accomplish this, then why, 15 years later, do we still lack an existence proof that a PDP model can do this?

To summarize the argument so far then, the Plaut and Booth (2000) model is problematic in two ways. The first problem is that there is as yet no evidence that this model can actually discriminate words from nonwords when the orthographic structure of the nonwords is the same as for the words. A corollary is that if it could make such discriminations, it remains to be demonstrated that it could also produce the same pattern of factor effects on RT that it claims to at present. A second problem is that there is no evidence that this or any other PDP model can produce lexical decision performance in the presence of severe semantic damage that is just as accurate as in the absence of semantic damage, yet, there exist multiple cases of patients with severe semantic damage

[1] Ironically, Borowsky and Besner (1993), Borowsky and Masson (1996), Stolz and Besner (1996, 1998), Brown and Besner (2002), Smith and Besner (2001), and Stolz and Neely (1995) also assumed that lexical decision is typically carried out at the semantic level so as to explain a variety of effects produced by intact university-level readers.

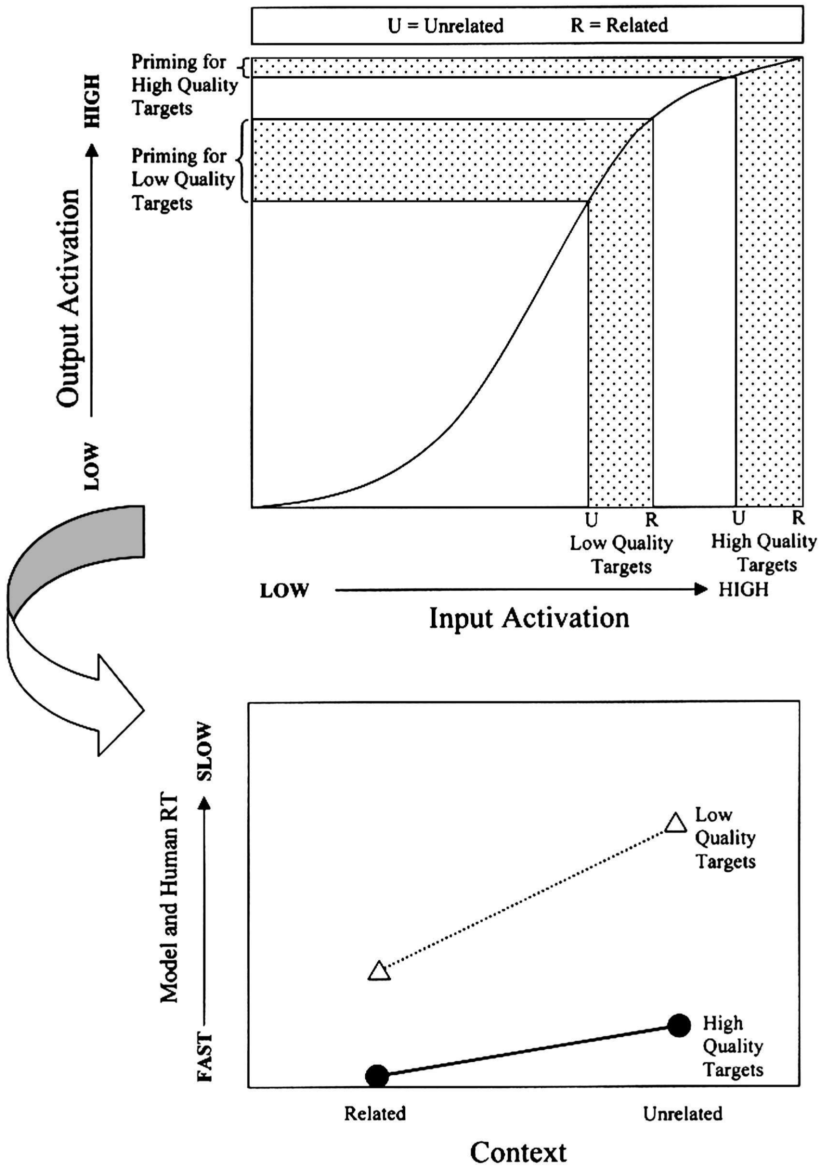

Figure 3. How the sigmoid activation function from the Plaut (1995) and Plaut and Booth (2000) model simulates the interaction between context and stimulus quality. RT = reaction time.

whose lexical decision performance is just as good as control patients without semantic damage (see Coltheart, 2004).

Problem three: Does the model successfully simulate the pattern of factor effects seen in the human literature? A critical feature of processing in Plaut and Booth's (2000) model is the shape of the activation function. Specifically, the output of the semantic system, which is based on the vector of activations of individual semantic features, is a sigmoid function of orthographic input to the semantic system (see Figure 2).

The claim is that in this PDP model, a sigmoid activation function makes it possible to account for the following: (a) the interaction between orthographic input strength and semantic context on RT, (b) the interaction between semantic context and word frequency on RT, and (c) additive effects of orthographic input strength and word frequency on RT.[2]

Plaut and Booth (2000) claim that the Plaut (1995) model can accommodate the additive and interactive effects of various factors reported by Borowsky and Besner (1993):

> "In addition to target frequency, stimulus quality also interacted with priming context (i.e., greater priming for degraded compared with

[2] To be clear, Plaut's (1995) model produced additive effects of stimulus quality and word frequency on priming RT difference scores, which does not mean that these two factors are additive on RT; only that the three-way interaction between stimulus quality, word frequency, and priming context on RT is not significant.

> intact stimuli), but stimulus quality did not interact with target frequency. Thus, the Plaut (1995) network exhibited the pattern of results found empirically by Borowsky and Besner (1993). . ." (Plaut & Booth, 2000, p. 789).

Plaut and Booth (2000) are also very clear in arguing that "This pattern of results makes it difficult—at least within an additive factors framework (Sternberg, 1969)—to locate context and frequency effects at the same stage of processing" (p. 787), and that with their present model they are "demonstrating that an implemented simulation that does not separate frequency and context effects and which lacks expectancy-based processes nonetheless reproduces the most important empirical findings" (p. 787).

How does the sigmoid activation function do this? Although Plaut and Booth (2000) claimed that the basis for an interaction on RT can be understood in terms of the nonlinear effects of the sigmoid activation function, they also acknowledged that there are limitations with their approach (which we will not reiterate here, see their p. 790, and General Discussion, pp. 809–817). Figures 3 and 4 (top panels) illustrate how Plaut's (1995) and Plaut and Booth's (2000) sigmoid activation function relates input strength to RT. The bottom panels of these figures show the RT effects from skilled readers that this network is claimed to simulate.

In Figure 3, the interaction between semantic context and stimulus quality (whereby the effect of context is larger for

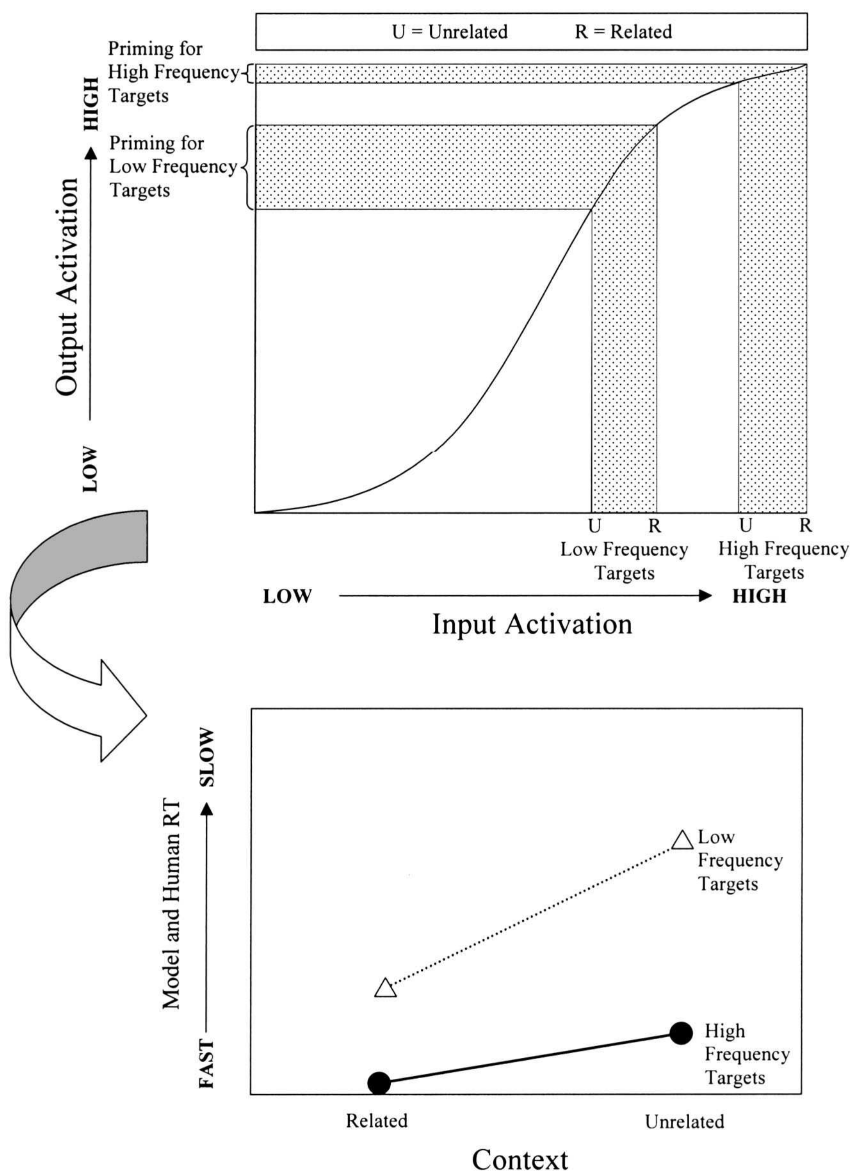

Figure 4. How the sigmoid activation function from the Plaut (1995) and Plaut and Booth (2000) model simulates the interaction between context and word frequency. RT = reaction time.

degraded rather than intact word targets) arises because the input strength to semantics from the orthographic units is stronger for words that follow a related prime than an unrelated one and stronger for intact targets than for degraded targets. The points at which the input strength values meet the sigmoid activation function translate into output (i.e., activation on the y-axis), which maps onto RT. The network RTs produce the pattern that is seen in human RTs as is illustrated in the bottom panel.

In Figure 4, the interaction between semantic context and word frequency (whereby the effect of context is larger for low-frequency rather than high-frequency word targets) arises because the input strength to semantics from the orthographic units is stronger for words that have followed a related prime

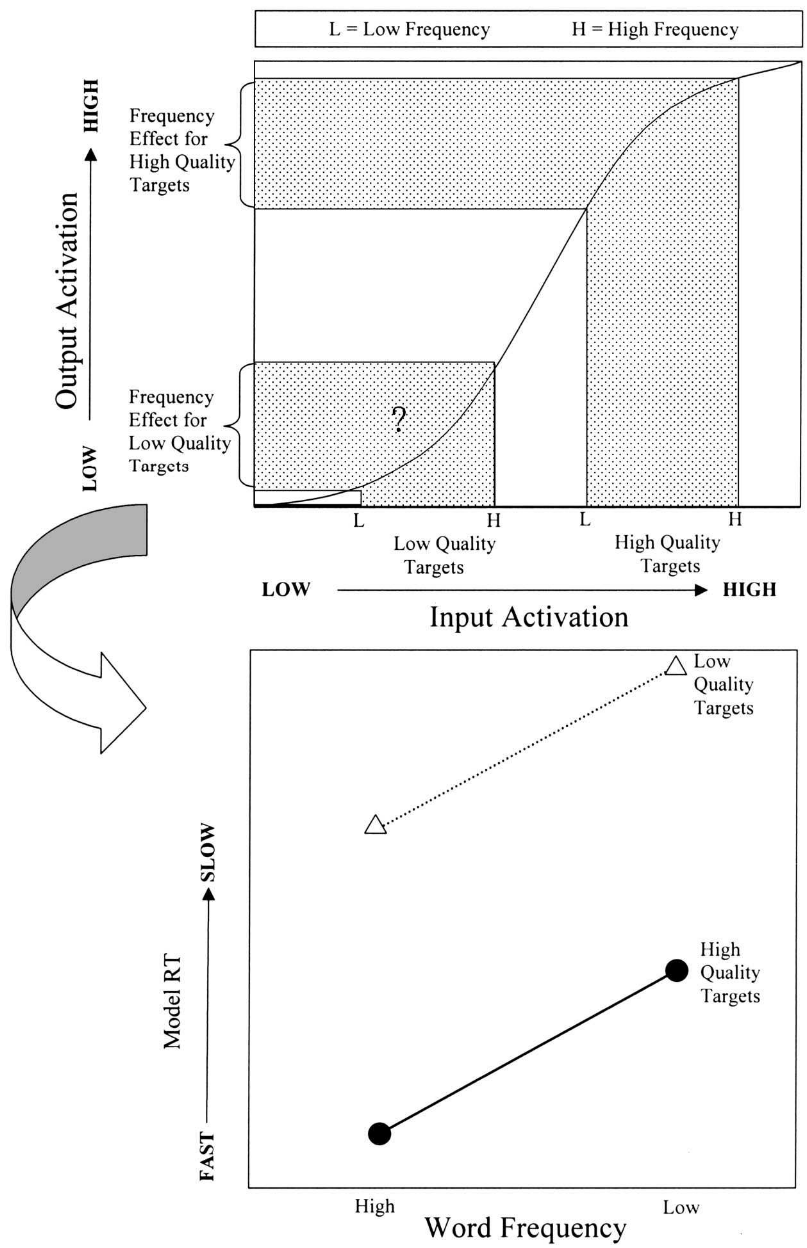

Figure 5. How the sigmoid activation function from the Plaut (1995) and Plaut and Booth (2000) model imposes some problematic constraints when trying to simulate the additive effects of stimulus quality and word frequency. The question mark represents the problematic region of the sigmoid activation function whereby some of the effects that are equidistant from the center of the sigmoid function impose unrealistic constraints on reaction time (RT). If one holds constant the high stimulus quality points from Figures 3 and 4, then the effect of stimulus quality is too large to accommodate the range of RTs of the interaction between context and word frequency (see Figure 4) and the interaction between context and stimulus quality (see Figure 3).

context than an unrelated one and stronger for high-frequency word targets than for low-frequency word targets. The points at which the input strength values meet the sigmoid activation function translate into network RTs that produce the pattern that is seen in human RTs as is illustrated in the bottom panel.

Figures 5 and 6 illustrate how the additive effects of stimulus quality and word frequency arise in the model. Again, points at which the input strength values meet the sigmoid activation function translate into network RTs that are supposed to produce the pattern that is seen in human RTs (see bottom panel). Critically, additive effects of these two factors are simulated by having the input strength of all four points of the two factors (stimulus quality and word frequency) meet the sigmoid activation function at points that are equidistant from the center (i.e., linear portion) of the function. Clearly, this model cannot simulate both an interaction and additive effects of factors as

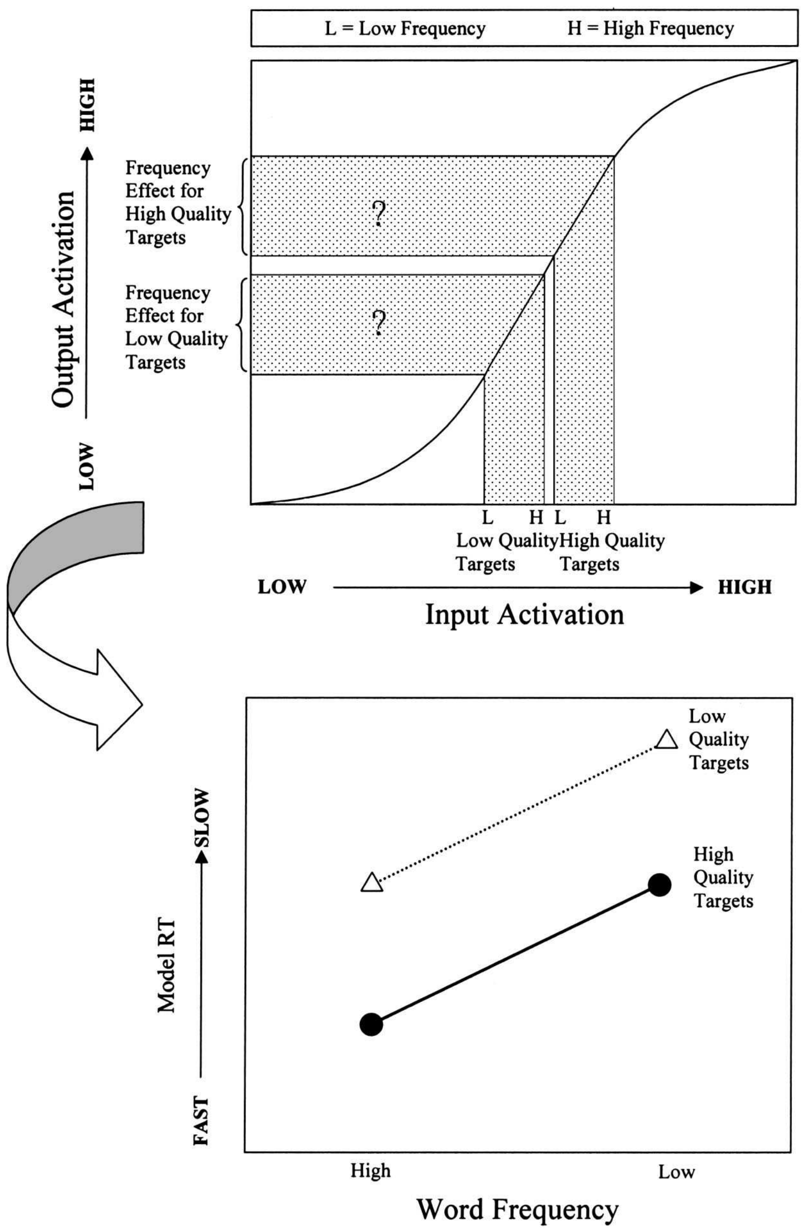

Figure 6. How the sigmoid activation function from the Plaut (1995) and Plaut and Booth (2000) model imposes some problematic constraints when trying to simulate the additive effects of stimulus quality and word frequency. The question marks represent the problematic region of the sigmoid activation function whereby some of the effects that are equidistant from the center of the sigmoid function impose unrealistic constraints on reaction time (RT). A more realistic size of effect of stimulus quality can be modeled by moving closer to the center of the sigmoid activation function (in contrast to Figure 5), but once again, at the cost of no longer accommodating the interactions in Figures 3 and 4.

described above when they all fall within the same range of RTs.

Figure 7 illustrates the general pattern of results when relying on a sigmoid activation function to simulate such effects, and thus illustrates how the sigmoid activation function, in and of itself, cannot accommodate the data when it is constrained within a range of RTs that reflects human performance (Borowsky & Besner, 1993; Brown & Besner, 2002; Stolz & Neely, 1995)

For example, Figure 8 illustrates the inherent problem of the sigmoid function if one were to attempt to account for additive effects of stimulus quality and word frequency within the same range of RTs as the interactive effects of context with stimulus quality (see Figure 3) and word frequency (see Figure 4). The sigmoid function incorrectly produces an interaction between orthographic input strength and word frequency. We now turn to a discussion of additional empirical findings that are inconsistent with the model.

The Context × Frequency Interaction Revisited

Plaut and Booth (2000) considered the context by frequency interaction on lexical decision RTs to have "important implications

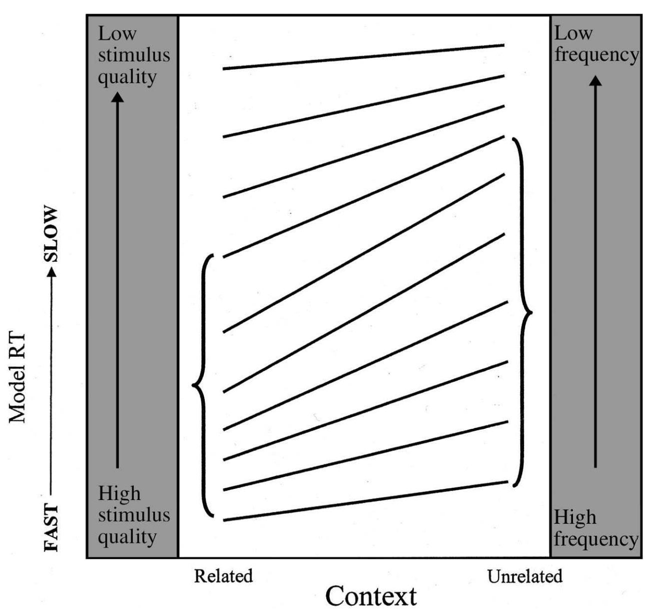

Figure 7. A depiction of how the sigmoid activation function from Plaut and Booth's (2000) simulation model generates effects of context, stimulus quality, and word frequency on reaction time (RT). The brackets capture the effects from the middle and upper regions of the sigmoid function, which Plaut & Booth (see their Figure 1) suggested could accommodate the interactive effects of context with word frequency and orthographic input strength (which also represents stimulus quality). This figure illustrates how a sigmoid activation function is not sufficient to accommodate both additive effects (e.g., stimulus quality and word frequency) and interactive effects (e.g., Context × Word Frequency, and Context × Stimulus Quality) within the same range of RTs. It further illustrates how the model produces a reverse interaction between context and word frequency if one examines only low stimulus quality targets, which is not observed in the human data (Borowsky & Besner, 1993). See Figure 9 for a more detailed depiction of this reverse interaction.

for theories of lexical processing" (p. 787) and thus chose to focus on this interaction throughout much of their paper (e.g., their Figure 1 illustrates this interaction). Plaut and Booth's model produces an interaction between context and word frequency (a larger effect of context for low-frequency words than high-frequency words under the high orthographic input strength condition), and this corresponds to what has been reported in the literature on university-level readers (Becker, 1979; Borowsky & Besner, 1993).

However, as seen in Figures 7 and 9, the model yields an interaction between context and word frequency that changes as a function of decreased orthographic input strength (recall that perceptual ability and stimulus quality are treated as functionally equivalent in this model, as both have been modeled by manipulations of orthographic input strength; Plaut, 1995; Plaut & Booth, 2000):

> "the network produced a trend toward a reverse Frequency × Context interaction (i.e., greater priming for high- compared with low-frequency targets) when tested in the adult, low-perceptual ability condition at the long SOA. The same pattern held numerically for these conditions at the short SOA . . . this reverse interaction, like the standard one in the high ability condition, can be understood in terms of the nonlinear effects of the sigmoid activation function (see Figure 1)." (Plaut & Booth, 2000, p. 810)

Although Plaut and Booth (2000) were cautious to not overinterpret this trend, the model actually produced more than a trend in that there were significant three-way interactions between perceptual ability, context, and word frequency in their simulations of adult lexical decision performance under both short and long prime-target stimulus onset asynchronies (SOA). The problem here is that this is not the pattern seen in Borowsky and Besner's (1993) data when stimulus quality is manipulated. There, the interaction between context and word frequency becomes larger rather than reversed under the low stimulus quality condition. The single-stage sigmoid activation function does not, therefore, correctly simulate these data.

Interactive and Additive Effects of Orthographic Input Strength and Semantic Priming

Plaut and Booth (2000) do not discuss the results of Stolz and Neely's (1995) Experiment 2, in which the SOA between prime and target was 200 ms. Stolz and Neely discussed these data at length and concluded that the results support the idea that conscious expectancies were not operating here. This is important because, in the absence of a conscious expectancy (which is outside the scope of the model), it places these findings squarely where the Plaut and Booth model should be able to account for them. Stolz and Neely replicated previous reports of an interaction between stimulus quality and semantic priming at a 200-ms SOA. Critically, however, that interaction was only seen when the proportion of related word trials was .50 and when the prime-target word pairs were strongly associated. The interaction was not seen when relatedness proportion was .25, and it was not seen when the strength of association between primes and targets was low. These data can be seen in Table 1. We note that the range of RTs across this experiment is from

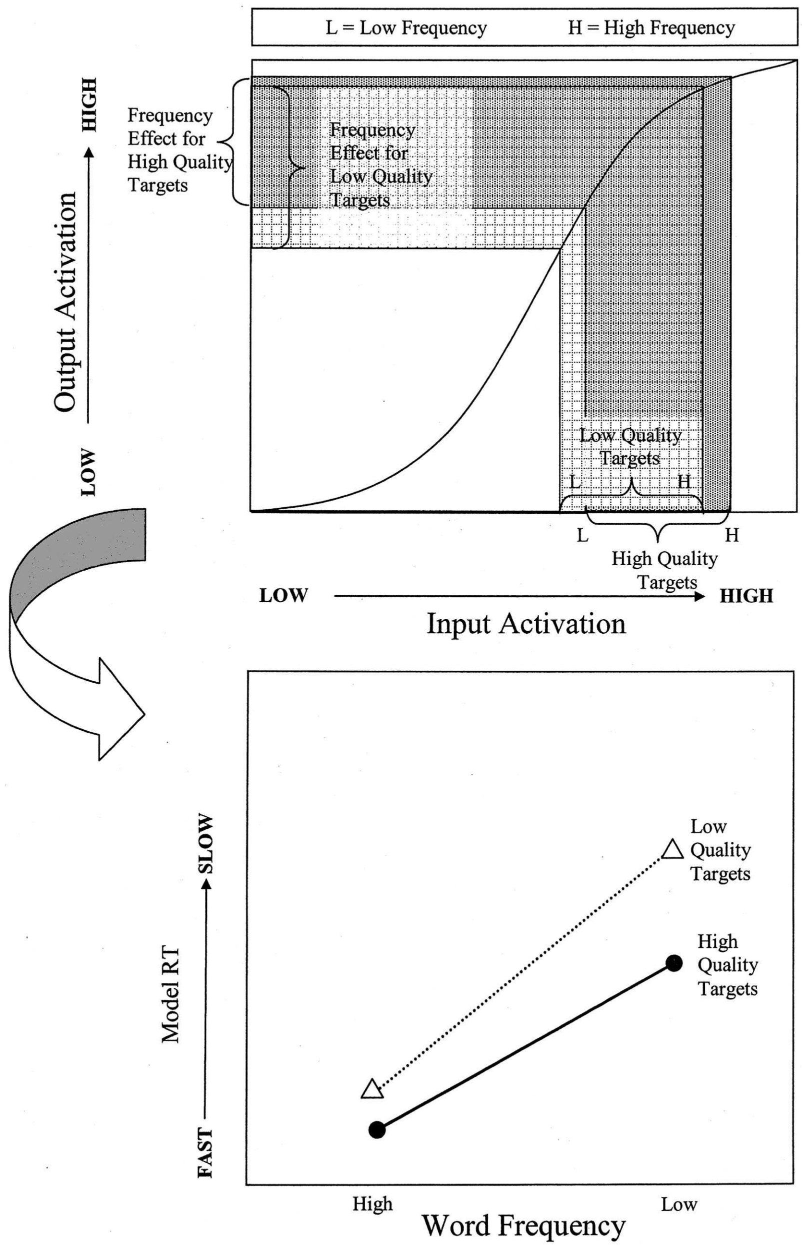

Figure 8. A depiction of how the sigmoid function (incorrectly) produces an interaction between orthographic input strength (stimulus quality) and word frequency if one attempts to maintain a similar range of reaction times (RTs) as are found in experiments with skilled readers that include all three variables (i.e., stimulus quality, word frequency, and semantic context, e.g., Borowsky & Besner, 1993).

596 ms (the bright, related condition with strong associates), to 754 ms (dim, unrelated condition with weak associates). Three of the four 2 × 2 cells in the matrix produce additive effects of stimulus quality and semantic priming. One 2 × 2 cell produces the interaction. Given that all of the 16 RT means fall within that range (approximately 600–750 ms), we do not see how one could appeal to the nonlinear sigmoidal activation function to explain both the interaction and the additive effects. In contrast, Stolz and Neely provided an explanation for these data (relatedly, see Brown & Besner, 2002), but it is couched in exactly the framework that Plaut and Booth wish to do without—multiple stages.

Revisiting How Stimulus Quality Affects Performance in the Model

Another concern is how a reduction in stimulus quality affects performance in the model. Figure 5 shows that the model

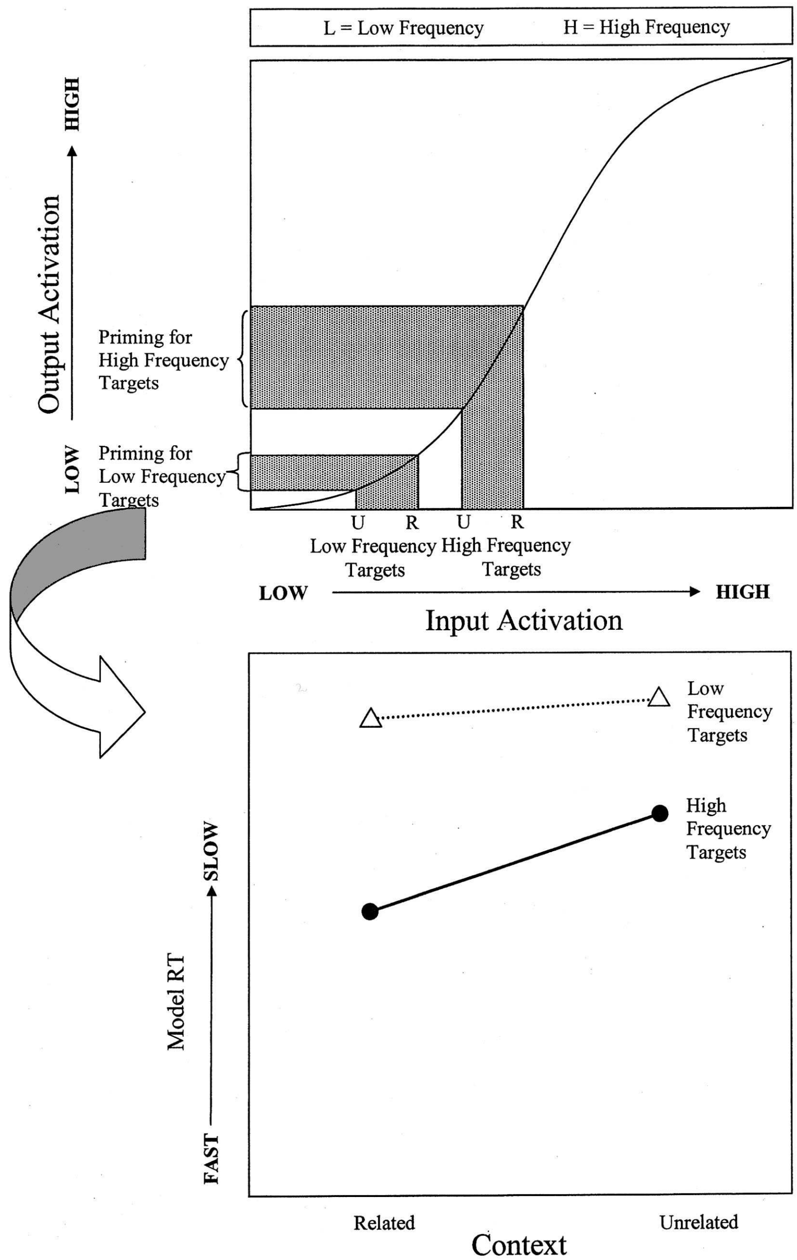

Figure 9. The model produces a Context × Frequency interaction in which low-frequency targets are less affected by context when orthographic input strength is reduced. Borowsky and Besner (1993) reported, for skilled participants, that low-frequency targets yielded a larger priming effect than high-frequency targets when stimulus quality is reduced. RT = reaction time.

can produce additive effects of stimulus quality and word frequency when the effects of both of these factors are restricted to the part of the activation function that is linear (or equidistant from this central portion). One reservation here is that the main effect of orthographic input strength produced by the simulation is tiny (approximately 12 ms when comparing the nonword prime baseline conditions—see Figure 14 in Plaut & Booth, 2000) when it was simulating the results of an experiment with adults in which the effect of perceptual ability was approximately 80–100 ms when comparing these same baseline conditions. Clearly, it is easy to produce a larger main effect of orthographic input strength in the model. The empirical question is whether the model is able to produce additive effects of two factors when the size of the orthographic input strength effect is equal to the perceptual ability effect that Plaut and Booth obtained with their experiments on humans, or similarly, equal to the stimulus quality effect obtained by others (e.g., Borowsky & Besner, 1993, 250–275 ms; Stolz & Neely, 1995,

Table 1
Mean RTs (ms) and Percentage Errors (%E) From Stolz and Neely's (1995) Experiment 2 (SOA = 200 ms) as a Function of RP, Relatedness, and Strength of Association

	RP = .25				RP = .50			
	Bright		Dim		Bright		Dim	
Associate	RT	%E	RT	%E	RT	%E	RT	%E
Strong								
Unrelated	630	1.6	726	2.3	628	2.4	754	2.8
Related	596	0.9	680	1.7	597	1.8	702	1.6
Difference	**34**	**0.7**	**46**	**0.6**	**31**	**0.6**	**52**	**1.2**
Weak								
Unrelated	640	2.9	746	3.9	637	2.2	754	4.1
Related	609	1.7	717	3.9	607	1.4	720	2.4
Difference	**31**	**1.2**	**29**	**1.5**	**30**	**0.8**	**34**	**1.7**

Note. RT = reaction time; SOA = stimulus onset asynchrony; RP = relatedness proportion.

100–125 ms). Unfortunately, for technical reasons, Plaut and Booth's model is not available, nor are the data from their simulations. (D. Plaut, personal communication, January 31, 2005). However, we understand that a replication is forthcoming.

Certainly, we already know from simulation work with an otherwise very successful model in which the activation function is monotonic (Coltheart et al.'s, 2001 dual route cascaded model) that reducing the rate of activation so as to simulate the effects of stimulus quality leads to an interaction with word frequency rather than additive effects (Reynolds & Besner, 2004).

Conclusions

The main conclusion reached here is that Plaut and Booth's (2000) single stage account of lexical processing imbedded in a PDP model does not advance our understanding of the processes underlying lexical decision by skilled readers. As we have discussed, the currently implemented version of the model fails in a number of fundamental ways. We therefore continue to prefer multiple stage accounts of such data (e.g., Borowsky & Besner, 1993; Brown & Besner, 2002; Smith & Besner, 2001; Stolz & Neely, 1995). What the functions of these stages are, how particular stages are represented (e.g., localist vs. distributed), whether these stages are cascaded or not, and how various factors constrain their operation, are issues that merit further empirical investigation and theoretical development.

References

Balota, D. A., & Abrams, R. A. (1995). Mental chronometry: Beyond onset latencies in the lexical decision task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 1289–1302.

Becker, C. A. (1979). Semantic context and word frequency effects in visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 5,* 252–259.

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., & Roberts, M. (2003). Reading nonwords aloud: Results requiring change in the dual route cascaded model. *Psychonomic Bulletin and Review, 10,* 398–401.

Besner, D., & Smith, M. C. (1992). Models of visual word recognition: When obscuring the stimulus yields a clearer view. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 468–482.

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

Borowsky, R., & Besner, D. (1991). Visual word recognition across orthographies: On the interaction between context and degradation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 272–276.

Borowsky, R., & Besner, D. (1993). Visual word recognition: A multistage activation model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 813–840.

Borowsky, R., & Masson, M. E. J. (1996). Semantic ambiguity effects in word identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 63–85.

Brown, M., & Besner, D. (2002). Semantic priming: On the role of awareness in visual word recognition in the absence of an expectancy. *Consciousness and Cognition, 11,* 402–422.

Carr, T. H., & Pollatsek, A. (1985). Recognizing printed words: A look at current models. In D. Besner, T. G Waller, & G. E MacKinnon. (Eds.), *Reading research: Advances in theory and practice* (pp. 2–82). Orlando, FL: Academic Press.

Coltheart, M. (2004). Are there lexicons? *Quarterly Journal of Experimental Psychology, 57A,* 1153–1171.

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., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance XI* (pp. 534–555). New York: Academic Press.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Zeigler, J. (2001). DRC: A dual route cascade model of visual word recognition and reading aloud. *Psychological Review, 108,* 204–256.

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, & Cognition, 10,* 680–698.

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

Joordens, S., & Besner, D. (1994). When banking on meaning is not (yet) money in the bank: Explorations in connectionist modeling. *Journal of*

Experimental Psychology: Learning, Memory, and Cognition, 20, 1051–1062.

Lambon Ralph, M. A., Ellis, A. W., & Franklin, S. (1995). Semantic loss without surface dyslexia. *Neurocase, 1,* 363–369.

Lambon Ralph, M. A., Sage, K., & Ellis. A. W. (1996). Word meaning blindness: A new form of acquired dyslexia. *Cognitive Neuropsychology, 13,* 617–639.

McClelland, J. L. (1979). On the time relations of mental processes: A framework for analyzing processes in cascade. *Psychological Review, 86,* 287–330.

McClelland, J. L. (1987). The case for interactionism in language processing. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 1–36). London: Erlbaum.

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

Meyer, D. E., Schvaneveldt, R. W., & Ruddy, M. G. (1975). Loci of contextual effects on visual word recognition. In P. M. A. Rabbitt & S. Dornic (Eds.), *Attention and performance V* (pp. 98–118). San Diego, CA: Academic Press.

Monsell, S. (1991). The nature and locus of word frequency effects in reading. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 148–197). Hillsdale, NJ: Erlbaum.

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

Morton, J. (1979). Facilitation in word recognition: Experiments causing change in the logogen model. In P. A. Kolers, M. E. Wrolstad, & H. Bouma (Eds.), *Processing of visible language* (Vol. 1, pp. 259–268). New York: Plenum Press.

Morton, J., & Patterson, K. (1980). A new attempt at an interpretation, or, an attempt at a new interpretation. In M. Coltheart, K. Patterson, & J. C. Marshall (Eds.), *Deep dyslexia* (pp. 91–118). London: Routledge & Kegan Paul.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General, 106,* 226–254.

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 264–336). Hillsdale, NJ: Erlbaum.

Norman, D. A. (1969). *Memory and attention*. New York: Wiley.

Plaut, D. C. (1995, July). Semantic and associative priming in a distributed attractor network. In *Proceedings of the 17th annual conference of the Cognitive Science Society* (pp. 37–42). Hillsdale, NJ: Erlbaum.

Plaut, D. C. (1997). Structure and function in the lexical system: Insights from distributed models of word reading and lexical decision. *Language and Cognitive Processes, 12,* 765–805.

Plaut, D. C., & Booth, J. R. (2000). Individual and developmental differences in semantic priming: Empirical and computational support for a single-mechanism account of lexical processing. *Psychological Review, 107,* 786–823.

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.

Plaut, D. C., & Shallice, T. (1993). Deep dyslexia: A case study of connectionist neuropsychology. *Cognitive Neuropsychology, 10,* 377–500.

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

Reynolds, M., & Besner, D. (2002). Neighborhood density effects in reading aloud: New insights from simulations with the DRC model. *Canadian Journal of Experimental Psychology, 56,* 310–318.

Reynolds, M., & Besner, D. (2004). Neigbourhood density, word frequency, and spelling-sound regularity effects in naming: Similarities and differences between skilled readers and the dual route cascaded computational model. *Canadian Journal of Experimental Psychology, 58,* 13–31.

Roberts, S., & Sternberg, S. (1993). The meaning of additive reaction-time effects: Tests of three alternatives. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 611–653). Cambridge, MA: MIT Press.

Rogers, T. T., Lambon Ralph, M. A., Hodges, J. R., & Patterson, K. (2004). Natural selection: The impact of semantic impairment on lexical and object decision. *Cognitive Neuropsychology, 21,* 331–352.

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

Shulman, H., Hornak, R., & Sanders, E. (1978). The effects of graphemic, phonemic, and semantic relationships on access to lexical structures. *Memory & Cognition, 6,* 115–123.

Smith, M. C., & Besner, D. (2001). Modulating semantic feedback in visual word recognition. *Psychonomic Bulletin and Review, 8,* 111–117.

Stanners, R. F., Jastrembski, J. E., & Westbrook, A. (1975). Frequency and visual quality in a word–nonword classification task. *Journal of Verbal Learning and Verbal Behavior, 14,* 259–264.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. In W. G. Koster (Ed.), *Attention and performance II* (pp. 276–315). Amsterdam: North-Holland.

Sternberg, S. (1998). Discovering mental processing stages: The method of additive factors. In D. Scarborough & S. Sternberg (Eds.), *Methods, models, and conceptual issues: An invitation to cognitive science* (Vol. 4, pp. 703–863). Cambridge, MA: The MIT Press.

Stolz, J. A., & Besner, D. (1996). The role of set in visual word recognition: Activation and activation blocking as non-automatic processes. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 1166–1177.

Stolz, J. A., & Besner, D. (1998). Levels of representation in visual word recognition: A dissociation between morphological and semantic processing. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1642–1655.

Stolz, J. A., & Neely, J. H. (1995). When target degradation does and does not enhance semantic context effects in word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 596–611.

Treisman, A. M. (1960). Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology, 12,* 242–248.

Ward, J., Stott, R., & Parkin, A. J. (2000). The role of semantics in reading and spelling: Evidence for the "summation hypothesis." *Neuropsychologia, 38,* 1643–1653.

Received April 2, 2004
Revision received March 2, 2005
Accepted March 3, 2005 ■

Postscript: Plaut and Booth's (2006) New Simulations—What Have We Learned?

Derek Besner
University of Waterloo

Ron Borowsky
University of Saskatchewan

Plaut and Booth's (2006) first simulation shows that there is essentially perfect discrimination between word and nonwords sharing the same orthographic structure when the simulation is carried out in the way we suggested. We are therefore satisfied that their model can (now) be characterized as a parallel distributed processing (PDP) model of lexical processing (with the caveat that although it is beyond the scope of the present work, it is important to show at some point that the model "scales up" when it has a more realistically sized vocabulary of 40,000 words or so).

We voiced concern that their model makes lexical decisions at the semantic level but there are patients with severe damage to semantics who are nonetheless as accurate at lexical decision as control patients without such damage (Coltheart, 2004). Can a model that makes lexical decisions at the semantic level simulate these data? In reply, Plaut and Booth (2006) reported new simulations showing that increasingly severe "lesioning" of their model at the semantic level impairs lexical decision accuracy[1] in a monotonic way, but it is a remarkably small effect (see their Figure 1). They argue that "distinguishing the semantic activation of one word from that of another requires far more detailed information—and, thus, is less robust to damage—than distinguishing either from the much weaker activation produced by a nonword" (Plaut & Booth, 2006, p. 198).

The results of these simulations not withstanding, a substantive difficulty for their account of lexical decision remains. Blazely, Coltheart, and Casey (in press) reported a detailed analysis of two patients (EM and PC) with semantic dementia, both of whom had significant impairments of semantic memory. EM performs slightly worse than PC on semantic tasks, but her visual lexical decision performance (two-alternative forced choice) was virtually perfect (97% correct), whereas PC's visual lexical decision performance was significantly impaired (75% correct).[2] It is difficult to see how this pattern can be simulated by Plaut and Booth's (2000) model if lexical decision is carried out at the semantic level, but it is easy to understand if the decisions are carried out at the lexical level and PC's lexical processing abilities are impaired. Blazely et al. provide converging evidence in support of this conclusion.

Our third major concern was that Plaut and Booth's (2000) PDP model's use of the sigmoid function relating activation to reaction time (RT) would not permit the joint effect of two factors on RT to be additive when one of these factors and a third factor produce an interaction that lies within the same range of RTs. They responded to this by reporting a new simulation that purports to accomplish this. These data are reported in their Table 1. Some of these data are reproduced in Figure 1 for illustrative purposes.

One point here is that this simulation is an unusual way of looking at the joint effects of word frequency and stimulus quality

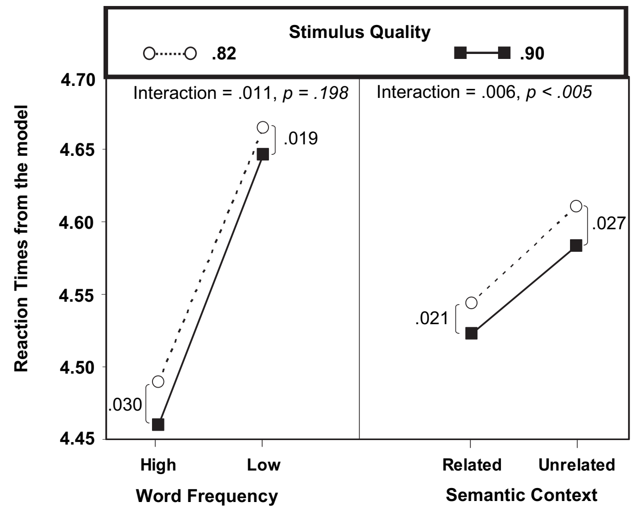

Figure 1. Reactions times from Plaut and Booth's (2006) model for the joint effects of stimulus quality and word frequency (left-hand panel) and stimulus quality and semantic priming (right-hand panel).

because it entails collapsing across the significant Context × Stimulus Quality interaction in the right-hand panel. The joint effects of stimulus quality and word frequency should be examined when there is a neutral prime rather than a related or unrelated word, as argued by Neely (1991; see also Borowsky & Besner, 1993, who used nonword primes, and see Plaut & Booth, 2000, who also used a nonword baseline in their experiments and simulations but have now dropped it in their 2006 simulation). The more important question is whether the model has produced additive effects of stimulus quality and word frequency (as claimed for the data in the left hand panel). The size of the nonsignificant (underadditive) interaction in the left panel (.011 units) is almost double the size of the significant interaction in the right-hand panel (.006 units). Plaut and Booth (2006) do not comment on this. Even if Plaut and Booth were able to provide a more convincing simulation of both additive and interactive effects within the same range of RTs, we would like to know what principled reason makes it possible for the simulation to produce these results given that they now agree that the sigmoid-based explanation "only approximates the actual behavior of the model" (Plaut & Booth, 2006, p. 199).

We take the view that Plaut and Booth's (2006) new simulation work settles little beyond the fact that their model can discriminate

[1] Plaut and Booth (2006) do not report what effect these lesions have on lexical decision RTs. Likewise, no published papers that we are aware of have reported lexical decision RTs for patients with severe semantic damage. It would be useful to have such data because they may help to further discriminate between alternative accounts.

[2] Patient PC's difficulties cannot be attributed to early deficits (e.g., letter identification) given that he reads regular words with high accuracy (94%).

between words and nonwords with the same orthographic structure. The idea that stages of processing underlie mental computation has a long history and has proven to be a useful theoretical framework for perception, cognition, action, and cognitive neuroscience. We see nothing here that persuades us it should be abandoned.

References

Blazely, A. M., Coltheart, M., & Casey, B. J. (in press). Semantic impairment with and without surface dyslexia: Implications for models of reading. *Cognitive Neuropsychology*.

Borowsky, R., & Besner, D. (1993). Visual word recognition: A multistage activation model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 813–840.

Coltheart, M. (2004). Are there lexicons? *Quarterly Journal of Experimental Psychology, 57A,* 1153–1171.

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading* (pp. 264–336). Hillsdale, NJ: Erlbaum.

Plaut, D. C., & Booth, J. R. (2000). Individual and developmental differences in semantic priming: Empirical and computational support for a single-mechanism account of lexical processing. *Psychological Review, 107,* 786–823.

Plaut, D. C., & Booth, J. R. (2006). More modeling but still no stages: Reply to Borowsky and Besner. *Psychological Review, 113,* 196–200.

United States Postal Service

Statement of Ownership, Management, and Circulation

1. Publication Title	2. Publication Number	3. Filing Date
Psychological Review	4 4 8 - 8 0 0	October 2005
4. Issue Frequency: Quarterly	5. Number of Issues Published Annually: 4	6. Annual Subscription Price: Mbr $65 Indiv $135 Inst $375

7. Complete Mailing Address of Known Office of Publication *(Not printer) (Street, city, county, state, and ZIP+4)*
750 First Street, N.E., Washington, D.C. 20002-4242
Contact Person: Barbara Spruill
Telephone: 202 336-5578

8. Complete Mailing Address of Headquarters or General Business Office of Publisher *(Not printer)*
750 First Street, N.E., Washington, D.C. 20002-4242

9. Full Names and Complete Mailing Addresses of Publisher, Editor, and Managing Editor *(Do not leave blank)*
Publisher *(Name and complete mailing address)*
American Psychological Association
750 First Street, N.E.
Washington, D.C. 20002-4242
Editor *(Name and complete mailing address)*
Keith Rayner, PhD, Dept. of Psychology
Tobin Hall, Univ. of Mass.
Amherst, MA 01003
Managing Editor *(Name and complete mailing address)*
Susan J.A. Harris
American Psychological Association
750 First Street, N.E., Washington, D.C. 20002-4242

10. Owner *(Do not leave blank. If the publication is owned by a corporation, give the name and address of the corporation immediately followed by the names and addresses of all stockholders owning or holding 1 percent or more of the total amount of stock. If not owned by a corporation, give the names and addresses of the individual owners. If owned by a partnership or other unincorporated firm, give its name and address as well as those of each individual owner. If the publication is published by a nonprofit organization, give its name and address.)*

Full Name	Complete Mailing Address
American Psychological Association	750 First Street, N.E.
	Washington, D.C. 20002-4242

11. Known Bondholders, Mortgagees, and Other Security Holders Owning or Holding 1 Percent or More of Total Amount of Bonds, Mortgages, or Other Securities. If none, check box → ☒ None

Full Name	Complete Mailing Address
NONE	

12. Tax Status *(For completion by nonprofit organizations authorized to mail at nonprofit rates) (Check one)*
The purpose, function, and nonprofit status of this organization and the exempt status for federal income tax purposes:
☒ Has Not Changed During Preceding 12 Months
☐ Has Changed During Preceding 12 Months *(Publisher must submit explanation of change with this statement)*

PS Form 3526, October 1999 *(See Instructions on Reverse)*

13. Publication Title: Psychological Review
14. Issue Date for Circulation Data Below: July 2005

15. Extent and Nature of Circulation		Average No. Copies Each Issue During Preceding 12 Months	No. Copies of Single Issue Published Nearest to Filing Date
a. Total Number of Copies *(Net press run)*		4700	4500
b. Paid and/or Requested Circulation	(1) Paid/Requested Outside-County Mail Subscriptions Stated on Form 3541. *(Include advertiser's proof and exchange copies)*	2873	2576
	(2) Paid In-County Subscriptions Stated on Form 3541 *(Include advertiser's proof and exchange copies)*		
	(3) Sales Through Dealers and Carriers, Street Vendors, Counter Sales, and Other Non-USPS Paid Distribution	830	845
	(4) Other Classes Mailed Through the USPS		
c. Total Paid and/or Requested Circulation *[Sum of 15b. (1), (2),(3),and (4)]*		3703	3421
d. Free Distribution by Mail *(Samples, compliment ary, and other free)*	(1) Outside-County as Stated on Form 3541	251	197
	(2) In-County as Stated on Form 3541		
	(3) Other Classes Mailed Through the USPS		
e. Free Distribution Outside the Mail *(Carriers or other means)*			
f. Total Free Distribution *(Sum of 15d. and 15e.)*		251	197
g. Total Distribution *(Sum of 15c. and 15f)*		3954	3618
h. Copies not Distributed		746	882
i. Total *(Sum of 15g. and h.)*		4700	4500
j. Percent Paid and/or Requested Circulation *(15c. divided by 15g. times 100)*		93.6	94.6

16. Publication of Statement of Ownership
☐ Publication required. Will be printed in the January 06 issue of this publication. ☐ Publication not required.

17. Signature and Title of Editor, Publisher, Business Manager, or Owner: [illegible], Sr. Director, Publishing Services — Date: 10/14/05

I certify that all information furnished on this form is true and complete. I understand that anyone who furnishes false or misleading information on this form or who omits material or information requested on the form may be subject to criminal sanctions (including fines and imprisonment) and/or civil sanctions (including civil penalties).

Instructions to Publishers

1. Complete and file one copy of this form with your postmaster annually on or before October 1. Keep a copy of the completed form for your records.
2. In cases where the stockholder or security holder is a trustee, include in items 10 and 11 the name of the person or corporation for whom the trustee is acting. Also include the names and addresses of individuals who are stockholders who own or hold 1 percent or more of the total amount of bonds, mortgages, or other securities of the publishing corporation. In item 11, if none, check the box. Use blank sheets if more space is required.
3. Be sure to furnish all circulation information called for in item 15. Free circulation must be shown in items 15d, e, and f.
4. Item 15h., Copies not Distributed, must include (1) newsstand copies originally stated on Form 3541, and returned to the publisher, (2) estimated returns from news agents, and (3), copies for office use, leftovers, spoiled, and all other copies not distributed.
5. If the publication had Periodicals authorization as a general or requester publication, this Statement of Ownership, Management, and Circulation must be published; it must be printed in any issue in October or, if the publication is not published during October, the first issue printed after October.
6. In item 16, indicate the date of the issue in which this Statement of Ownership will be published.
7. Item 17 must be signed.

Failure to file or publish a statement of ownership may lead to suspension of Periodicals authorization.

PS Form 3526, October 1999 *(Reverse)*