Psychological Review
2006, Vol. 113, No. 1, 196–200

0033-295X/06/$12.00 DOI: 10.1037/0033-295X.113.1.196

More Modeling but Still No Stages: Reply to Borowsky and Besner

David C. Plaut
Carnegie Mellon University

James R. Booth
Northwestern University

Plaut and Booth (2000) developed a distributed connectionist model of written word comprehension and evaluated it against empirical findings on individual and developmental differences in semantic priming in visual lexical decision. Borowsky and Besner (2006) raised a number of challenges for this model. First, the model was not shown to be capable of accurately distinguishing words from orthographically matched nonwords. Second, its use of a semantic measure for performing lexical decision appears inconsistent with evidence of normal lexical decision in brain-damaged patients with semantic impairments. Third, the explanation offered for additive and interactive effects in the model appears incompatible with certain aspects of existing empirical findings on the joint effects word frequency, priming context, and stimulus quality. In this reply, the authors demonstrate with additional modeling that none of these issues is problematic for the model.

Key words: connectionist modeling, lexical processing, semantic priming

The study of semantic priming in lexical tasks has yielded rich and highly intricate patterns of results that have provided important insights into the nature of lexical and semantic processing (see McNamara, 2005; Neely, 1991, for reviews). As various theories have grappled with the complexities of these findings, though, they have been led to augment basic processing mechanisms (e.g., spreading activation) with additional, seemingly independent mechanisms (e.g., expectancy-list generation, retrospective matching, resource limitations), many of which seem directly tailored to the experimental priming manipulation itself. In this way, researchers have begun offering theories of priming rather than theories of lexical processing that give rise to priming.

In Plaut and Booth (2000), we offered an account of lexical processing, grounded in general (connectionist) principles, that we claimed could account for certain aspects of the relevant empirical phenomena, including some findings thought to implicate additional mechanisms. We supported this account by a specific computational implementation that, although necessarily limited and intentionally simplified, instantiated the core aspects of the theory. To enable the quantitative adequacy of the implemented model to be evaluated in detail, we developed it in the context of three specific empirical studies: adults and children tested on primed lexical decision at a long (800 ms) stimulus onset asynchrony (SOA) and adults tested at a short (200 ms) SOA. Previous work had found greater semantic priming (i.e., faster reaction times [RTs] following related vs. unrelated primes) for low- versus high-frequency target words and inhibition (i.e., slower RTs following unrelated vs. neutral primes) at a long, but not short, SOA (see Neely, 1991). We examined the extent to which these effects depended on individual differences among participants in age or perceptual ability. In brief, we found that greater priming for low-frequency targets was exhibited only by participants with high perceptual ability and that adults but not children exhibited inhibition at the long SOA. We went on to show that our implemented model behaved similarly. Here, we present a brief review of the Plaut and Booth model before addressing recent challenges to it.

David C. Plaut, Department of Psychology, Carnegie Mellon University; James R. Booth, Department of Communication Sciences and Disorders, Northwestern University.

Financial support was provided by National Institutes of Health Grant MH55628. A running version of the simulation can be downloaded from http://www.cnbc.cmu.edu/~plaut/xerion/

Correspondence concerning this article should be addressed to David C. Plaut, Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213-3890. E-mail: plaut@cmu.edu

The Plaut and Booth (2000) Model

The Plaut and Booth (2000) model is a fully recurrent, distributed connectionist model trained on an abstract task analogous to written word comprehension. Input patterns for 128 words were encoded over three banks of six units, each of which coded one of 15 letters by two active units. Five of the two-unit patterns ("vowels") occurred only in the middle bank; the remaining 10 ("consonants") occurred only in the other two banks. In this way, the input patterns could be construed as having a consonant–vowel–consonant (CVC) structure, although it should be clear that the degree to which the representations captured real orthographic structure—even among CVC words—was minimal.

Output patterns were created by first generating eight random binary prototype vectors of 100 elements and then randomly distorting them to create 16 exemplars from each prototype vector (half of which were *high-dominance* in that they involved less distortion than the remaining ones). In this way, the output patterns could be construed as being organized into semantic categories (i.e., clusters of patterns with relatively high feature overlap) although, again, the degree to which the representations captured real semantic structure was minimal. The output patterns were assigned randomly to input patterns to instantiate the critical property that, among monomorphemic words, orthographic similarity is unrelated to semantic similarity.

The model was trained with back-propagation (Rumelhart, Hinton, & Williams, 1986) to activate the semantic output features of each of the 128 words when presented with its orthographic input features. Input was presented to the network, not by fixing the states of input units, but by driving them with external input; the strength of this external input was used to approximate individual differences in perceptual ability. In processing a given word, the network started from the final pattern produced by the preceding word, and words were sampled such that those from the same semantic category (i.e., those generated from the same prototype) tended to follow each other during training. In addition, half of the words in each category were presented twice as often as the others.

At various points in training (corresponding to age differences), the model was tested on lexical decision using prime-target pairs of words and nonwords (i.e., novel orthographic patterns). The prime stimulus was presented and processed for a relatively short or long amount of time (corresponding to SOA) and then replaced by the target stimulus, which was processed until the output activations stabilized. Lexical decisions were based on a measure, termed *stress*, of the degree to which the stable semantic activations were close to binary and the processing time required to reach stability for the target was taken as the model's response latency. Thus, the model's accuracy and latency could be evaluated as a function of six factors: word frequency (number of training presentations), category dominance (amount of distortion from category prototype), priming context (related vs. unrelated vs. nonword primes), SOA (short vs. long prime processing), age (total amount of training), and perceptual ability (strength of external input).

The Plaut and Booth (2000) model is an example of what Kello and Plaut (2003) termed a *fundamentalist* approach to cognitive modeling. Instead of attempting to incorporate realistic scale and complexity in the task it addresses, it abstracts away as much extraneous detail as possible, embodying only those principles and properties that are claimed to account for the relevant phenomena. Although this approach has important limitations, it can provide a clearer understanding of why the principles give rise to the phenomena. Plaut and Booth attempted to contribute to this goal by explaining the model's performance by reference to properties of the sigmoid activation function used by output units (and others) in the network and how the various factors contribute to the strength of input provided to these units by the rest of the network.

Although a fundamentalist model is, by its very nature, highly abstract, it must nonetheless be evaluated in detail to determine the extent to which its behavior provides support for the underlying theory. Borowsky and Besner (2006) challenged the empirical adequacy of the Plaut and Booth (2000) implementation and hence the degree to which it supports our theoretical claims that a single (connectionist) mechanism suffices for lexical processing. They offered three main criticisms: (a) the model's ability to accurately distinguish words from orthographically matched nonwords was not established; (b) the model's reliance on semantics to perform lexical decision is contradicted by normal lexical decision accuracy in patients with semantic impairments; and (c) the model fails to account for the empirically observed relationships among the effects of word frequency, priming context, and stimulus quality. We take up each of these criticisms in turn.

Lexical Decision With Orthographically Matched Nonwords

All three of Plaut and Booth's (2000) empirical studies used essentially the same stimulus materials. Analyses indicated that the words and nonwords were not matched orthographically—for example, the mean summed positional bigram frequency for nonwords was only 76% that of words (62.4 vs. 82.0, respectively). Accordingly, to provide an appropriate comparison to the empirical data, the Plaut and Booth model was tested using nonwords with approximately the same degree of similarity to the artificial words as held among the experimental stimuli. Given the limitations of the available orthographic forms, the only way to accomplish this was to reverse the distribution of consonants and vowels—that is, to use nonwords with vowel–consonant–vowel (VCV) structure. The resulting nonwords had an average of 1.87 features in common with words, which is 84% that of the mean word–word overlap of 2.22 features. Thus, the similarity between words and nonwords was, if anything, slightly greater in the simulation than in the empirical studies.

Borowsky and Besner (2006) pointed out that it would be trivial for human participants to distinguish real VCV nonwords from CVC words in an empirical study. This is, of course, true but bears little relevance for evaluating the simulation, which used neither real words nor nonwords. More to the point, they also questioned whether the model can, like human participants, accurately perform lexical decision on the basis of semantic stress when nonwords are matched orthographically to words. Although this has already been demonstrated with real words and nonwords in a more full-scale distributed connectionist model (Plaut, 1997), it is worth evaluating in the current context.

Accordingly, a replication of the Plaut and Booth (2000) simulation[1] was tested after 200,000 word presentations—equivalent to Plaut and Booth's adult condition—for its ability to distinguish the 128 trained CVC words from the remaining 372 CVC orthographic forms as nonwords. For comparison, performance was also measured on the original 128 VCV nonwords. Each word or nonword was presented both in isolation and preceded by every word as prime at both the long and short SOAs.

When stimuli were presented in isolation, the distributions of stress values for words and nonwords did not overlap; stress values for words ranged from 0.943 to 0.978, whereas they ranged from 0.793 to 0.930 for CVC nonwords and from 0.774 to 0.916 for VCV nonwords. When stimuli were presented following word primes, a stress threshold of 0.927 yielded 99.9% hits, 99.9% correct rejections of CVC nonwords, and 100% correct rejections of VCV nonwords. Thus, the model is essentially perfect at distinguishing words from orthographically matched nonwords.

Lexical Decision With Semantic Impairment

As just described, lexical decisions by the Plaut and Booth (2000) model are based on semantic stress. This design decision

[1] The original simulation reported by Plaut and Booth (2000) was lost because of a computer malfunction and faulty backup storage. The replication reported here followed all of the simulation methods reported in the original article. It is not exactly equivalent because of differences in the initialization of the pseudorandom number generator and minor changes in the simulator code made subsequent to the original work.

was more for practical than theoretical reasons, as we were careful to point out (p. 812):

> Other network models and empirical findings have illustrated the importance of examining the interaction among orthographic, phonological, and semantic representations when trying to account for behavioral data from naming and lexical decision tasks. . . . Thus, the current simulation, which involved only a mapping from orthography to semantics, cannot be expected to provide a full account of lexical processing in general, nor even of lexical decision performance in particular. . . . In a more comprehensive version of our account of lexical decision, we would assume that subjects can base their decisions on any available information in the lexical system, and that they adopt a strategy that optimizes their performance given the composition of the stimuli. . . .

Nonetheless, Borowsky and Besner (2006) objected to the use of a semantic measure for making lexical decisions on the basis of evidence reviewed by Coltheart (2004) that some brain-damaged patients with semantic impairments exhibit normal lexical decision accuracy.

Without necessarily agreeing that all of Coltheart's (2004) evidence should be taken at face value, it is still worth examining how lexical decision performance in the model is influenced by semantic damage. Semantic lesions were administered to the network by selecting a specified proportion of semantic units at random and fixing their activations at zero. Performance on lexical decision to isolated words and CVC nonwords was determined following 32 instances of lesion at each of five levels of severity—0.025, 0.05, 0.1, 0.2, and 0.3—in which the decision criterion was set separately for each severity to yield a beta value near 1.0. For comparison, semantic performance on each word was measured in terms of whether the network succeeded in activating its semantic representation fully accurately (thresholding unit activations at 0.5).[2] Figure 1 shows the lexical decision and semantic performance of the model as a function of the proportion of semantic units lesioned. The results demonstrate that lexical decision performance is relatively unaffected over a range of lesion severities that produce substantial semantic impairment. The reason is, essentially, 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.

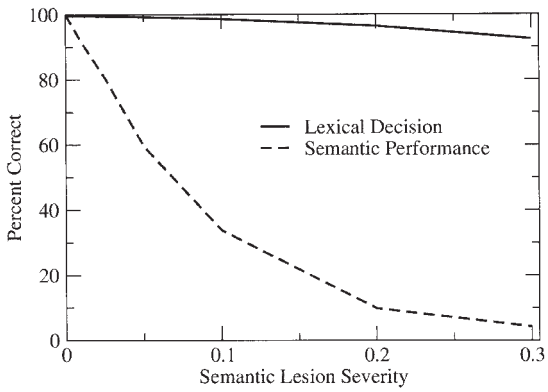

Figure 1. Correct performance of the model on lexical decision and on a measure of semantic performance as a function of the proportion of semantic units lesioned.

Effects of Word Frequency, Priming Context, and Stimulus Quality

The most extensive of Borowsky and Besner's (2006) criticisms concerned whether the Plaut and Booth (2000) model can account for the joint effects of word frequency, priming context, and stimulus quality. Although Plaut and Booth did not explicitly address effects of stimulus quality, an earlier, similar model (Plaut, 1995) did, using the same computational manipulation of external input strength that Plaut and Booth used to approximate individual differences in perceptual ability. Indeed, the earlier model exhibited the empirically observed qualitative pattern of performance—priming context interacted with both word frequency and stimulus quality, but the latter two did not interact.[3] Given the similarities of the two models, it seems reasonable to examine whether the account offered by Plaut and Booth for perceptual ability would be expected to generalize to stimulus quality. It should be kept in mind, though, that there are presumably many other ways of modeling variations in stimulus quality (e.g., reducing contrast, adding noise) and nothing in Plaut and Booth (2000) depends on a claim that manipulating external input strength per se is the best approach.

Borowsky and Besner's (2006) analysis of these issues was based not on the actual behavior of the Plaut and Booth (2000) model, but on their interpretation of a sigmoid diagram that we used to explain additive and interactive effects in the model. In brief, Plaut and Booth explained interactions in terms of the diminishing returns of the nonlinear (sigmoid) unit activation function—when one factor is sufficiently strong to drive activations into the asymptotic range of the sigmoid function, the effect of a second factor will be diminished. In contrast, when both factors fall within the linear range of the function (or equidistant from the center), the factors will combine additively. Borowsky and Besner argued that this account precludes additive and interactive patterns within the same RT range. In particular, they discussed three sets of empirical findings that appear problematic for the account.

Borowsky and Besner (1993)

Although the high-perceptual-ability condition in the Plaut and Booth (2000) model produced the standard finding of greater priming for low- versus high-frequency words, the low-perceptual-

[2] This measure of semantic performance is, of course, only indirectly related to observable performance on behavioral tasks thought to tap conceptual knowledge. Tasks for which less than fully accurate semantic representations suffice would be expected to show greater preservation with mild and moderate damage, and lexical decision performance degrades with more severe semantic damage (as observed empirically by Rogers, Lambon Ralph, Hodges, & Patterson, 2004).

[3] Borowsky and Besner (2006) point out that Plaut (1995) reported the lack of an interaction of frequency and stimulus quality only in terms of difference scores (unrelated minus related priming contexts). However, the factors also did not interact in the model's base RTs, $F(3, 378) = 0.31$, $p = .992$.

ability condition exhibited a slight tendency toward the reverse pattern numerically, although the effect was never statistically reliable. By contrast, Borowsky and Besner (1993) found empirically that their low-stimulus-quality condition produced not the reverse pattern, but a stronger version of the standard pattern.

Note, however, that Plaut and Booth's (2000) low-perceptual-ability participants behaved like the model in that they also produced a numerical (but not reliable) reverse interaction. Thus, under the assumption that stimulus quality and perceptual ability are equivalent, there is an empirical discrepancy between Borowsky and Besner's (1993) findings and those of Plaut and Booth. This may be one indication that using the same manipulation for both factors is inappropriate.

There is, however, a way to reconcile the two sets of findings using the properties of the sigmoid function. Perhaps Borowsky and Besner's (1993) low-stimulus-quality condition was analogous to Plaut and Booth's (2000) high-perceptual-ability condition, falling in the upper range of the sigmoid that produces the standard interaction. The low-perceptual-ability condition would then fall in the linear range or perhaps even slightly into the lower half of the sigmoid, as Plaut and Booth suggest. By contrast, Borowsky and Besner's high-stimulus-quality condition would fall even further toward asymptote, in which the frequency-by-context interaction is reduced because of a ceiling effect. Although this proposal is speculative, it suggests that Borowsky and Besner's findings can be accommodated within our account.

Stolz and Neely (1995)

Borowsky and Besner (2006) discussed Experiment 2 of Stolz and Neely (1995), which measured how the influence of stimulus quality on semantic priming depends on relatedness proportion and association strength. Although all four combinations of these factors yielded comparable RT ranges, only the combination of high association strength and high relatedness proportion yielded a reliable interaction of stimulus quality and priming context.

An initial point to make is that the data exhibit the general trend that the interaction becomes stronger as each individual factor increases in strength. Although it is reliable only in the case in which both relatedness proportion and association strength are high, the case in which only association strength is high produced 12 ms more priming for low- versus high-quality targets, although the difference wasn't quite reliable, $t(190) = 1.40$, $p = .0816$, one-tailed (misreported by Stolz & Neely, 1995, p. 605, as $p > .10$). This general pattern across conditions is consistent with Plaut and Booth's (2000) account of additive and interactive effects.

Second, it should be noted that the manipulation of relatedness proportion in the experiment involved different sets of participants, and the manipulation of association strength used different target words and was blocked within the experiment. Thus, rather different response criteria may have been operating across the four combinations of relatedness proportion and association strength, making any comparison of absolute RT ranges difficult to interpret.

However, even if the comparisons of RT ranges were accepted, it turns out that the resulting pattern is not problematic for the model. Collapsing over both long and short SOAs and considering only word primes, the model shows reliable main effects of word frequency, $F(1, 126) = 77.53$, $p < .001$, stimulus quality (0.9 vs. 0.82 as in Plaut & Booth, 2000), $F(1, 126) = 34.99$, $p < .001$, and priming context, $F(1, 126) = 202.9$, $p < .001$, reliable interactions of context with frequency, $F(1, 126) = 10.04$, $p < .005$, and context with quality, $F(1, 126) = 9.28$, $p < .005$, but no interaction of frequency and quality, $F(1, 126) = 1.67$, $p = .198$.[4] Table 1 lists the mean settling times for the model for each combination of factors. Averaging the relevant table entries reveals that the range of mean settling times exhibiting the interaction of context and frequency (4.450–4.694) contains the ranges that produce both the interaction of context and quality (4.523–4.611) and the additive combination of frequency and quality (4.460–4.665). Thus, the model exhibits both interactions and additive effects within the same RT range. Insofar as this pattern of results is precluded by Plaut and Booth's sigmoid-based explanation of additive and interactive effects—as argued by Borowsky and Besner (2006)—the results suggest that the account only approximates the actual behavior of the model.

Table 1
Settling Times for the Model as a Function of Word Frequency, Priming Context (Related vs. Unrelated), and Stimulus Quality

Stimulus quality	High frequency		Low frequency	
	Related	Unrelated	Related	Unrelated
0.9	4.437	4.483	4.609	4.684
0.82	4.463	4.516	4.625	4.705
0.75	4.497	4.555	4.658	4.749

Magnitude of Stimulus Quality Effects

A final criticism raised by Borowsky and Besner (2006) is that the magnitude of effects of stimulus quality (or, rather, perceptual ability) exhibited by the model is smaller than observed empirically. Although the specific numbers they cited are not representative, the general point holds: Across all adult conditions in Plaut and Booth (2000), the mean effect of perceptual ability was 35.3 ms for the model but 80.0 ms for human participants (i.e., 2.27 times larger). Borowsky and Besner expressed concern that a stronger manipulation of external input strength would compromise the additive relationship of stimulus quality and word frequency.

When the model is tested using a broader range of external input strength (0.90 vs. 0.75), the magnitude of the effect of stimulus quality increases by a factor of 2.54 (see Table 1). Nonetheless, the

[4] The interaction of stimulus quality and priming context in the model depends on SOA, $F(1, 126) = 6.41$, $p < .05$, such that it is reliable at the long SOA, $F(1, 126) = 10.84$, $p < .001$, but not at the short SOA, $F(1, 126) = 1.31$, $p = .255$. The same pattern holds when using a stronger manipulation of input strength (0.75 vs. 0.90). Similarly, Stolz and Neely (1995) found a reliable quality-by-context interaction at a 800-ms SOA (Experiment 1) but not at a 200-ms SOA (Experiment 2) when collapsing across relatedness proportion and association strength, although planned *t* tests for the short SOA showed an overadditive effect for strong associates under a high relatedness proportion. The model cannot be evaluated against this latter finding because the relevant factors were not included in the simulation.

basic qualitative pattern of performance remains unchanged: main effects of word frequency, $F(1, 126) = 72.36$, $p < .001$, priming context, $F(1, 126) = 189.7$, $p < .001$, and stimulus quality, $F(1, 126) = 92.86$, $p < .001$, interactions of context with frequency, $F(1, 126) = 9.90$, $p < .005$, and with quality, $F(1, 126) = 9.41$, $p < .005$, but, critically, still no interaction of frequency and quality, $F(1, 126) = 0.401$, $p = .528$.

Conclusion

Borowsky and Besner (2006) raised a number of challenges for Plaut and Booth's (2000) implemented model of semantic priming effects in lexical decision, calling into question the degree to which it provides support for a single-mechanism account of lexical processing. First, can the model discriminate words from orthographically matched nonwords? Second, is the use of a semantic measure for performing lexical decision compatible with preserved lexical decision in patients with semantic impairments? Third, is the model consistent with existing findings on interactive and additive effects among word frequency, priming context, and stimulus quality?

The answer to these questions seems to be yes. Despite the fact that the model is, by design, highly abstract in its instantiation of lexical processing, it nonetheless appears to capture important properties of human lexical processing. In light of these positive findings, we find no compelling reason to adopt Borowsky and Besner's (2006) suggestion that theories of lexical processing stipulate multiple explicit stages of processing.

References

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., & Besner, D. (2006). Parallel distributed processing and lexical-semantic effects in visual word recognition: Are a few stages necessary? *Psychological Review, 113,* 181–195.

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

Kello, C. T., & Plaut, D. C. (2003). Strategic control over rate of processing in word reading: A computational investigation. *Journal of Memory and Language, 48,* 207–232.

McNamara, T. S. (2005). *Semantic priming: Perspectives from memory and word recognition.* New York: Psychology Press.

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. (1995). Semantic and associative priming in a distributed attractor network. In J. D. Moore & J. F. Lehman (Eds.), *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 naming and lexical decision. *Language and Cognitive Processes, 12,* 767–808.

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.

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.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986, October 9). Learning representations by back-propagating errors. *Nature, 323,* 533–536.

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.

Received April 19, 2005
Revision received June 22, 2005
Accepted June 23, 2005 ■