Journal of Experimental Psychology:
Human Perception and Performance
2008, Vol. 34, No. 1, 242–250

0096-1523/08/$12.00 DOI: 10.1037/0096-1523.34.1.242

Constraints on Computational Models of Basic Processes in Reading

Derek Besner, Szymon Wartak, and Serje Robidoux
University of Waterloo

There are numerous reports in the visual word recognition literature that the joint effects of various factors are additive on reaction time. A central claim by D. C. Plaut and J. R. Booth (2000, 2006) is that their parallel distributed processing model simulates additive effects of stimulus quality and word frequency in the context of lexical decision. If correct, this success would have important implications for computational accounts of reading processes. However, the results of further simulations with this model undermine this claim given that the joint effects of stimulus quality and word frequency yield a nonmonotonic function (underadditivity, additivity, and overadditivity) depending on the size of the stimulus quality effect, whereas skilled readers yield additivity more broadly. The implications of these results both locally and more globally are discussed, and a number of other issues are noted. Additivity of factor effects constitutes a benchmark that computational accounts should strive to meet.

Keywords: computer simulation, word recognition, semantic priming, word frequency, lexical decision

It is widely accepted that mental computation in various reading tasks involves a number of distinct special purpose modules. A widely entertained theoretical account assumes that processing in one module feeds activation forward to other modules as soon as activation begins in the original module (processing is cascaded) and is accompanied by feedback between adjacent modules. These assumptions appear in models with very different types of representations (e.g., localist models vs. parallel distributed processing [PDP] models; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Grainger & Jacobs, 1996; Harm & Seidenberg, 2004; McClelland, 1987, 1991; McClelland & Rumelhart, 1981; Perry, Ziegler, & Zorzi, 2007; Plaut, McClelland, Seidenberg, & Patterson, 1996; Rogers & McClelland, 2004).

It is not difficult to intuit how cascaded activation across discrete representational levels combined with feedback between some of these levels yields interactions between various experimental factors. It is less obvious that such processing can also produce additive effects of two factors on reaction time (RT). This question is of interest given that there are numerous reports of such additivity in the visual word recognition literature on skilled readers over the past 30 years (e.g., see Table 1 for a number of examples).

To date, the issue of additivity of factor effects has been addressed by few computational modelers working on visual word recognition.[1] A singular exception is Plaut and Booth (2000), who asserted that their PDP model simulates additive effects of two particular factors on RT. This claim, among others, has been the source of some debate between Borowsky and Besner (2006; Besner & Borowsky, 2006) and Plaut and Booth (2006). A central aspect of this ongoing debate has not been resolved and is therefore revisited here.

We first briefly describe the highlights of the Plaut and Booth (2000) model. We then revisit a central issue raised by Borowsky and Besner (2006) and the simulation data reported by Plaut and Booth (2006) in their response. We then reconsider this issue with a set of new simulations. We (a) conclude that the Plaut and Booth (2000) model does not simulate the pattern observed with skilled readers, (b) discuss some reasons why this might be so, (c) briefly discuss some other examples of additivity in the literature that are not restricted to lexical decision, and (d) argue that computational models of reading processes should view additivity of factor effects as a basic benchmark.

Computational Models and Lexical Decision

To the best of our knowledge, only a few computational models have been developed that purport to explain a subset of results from the vast lexical decision literature (Seidenberg & McClelland, 1989—but see Besner, Twilley, McCann, & Seergobin, 1990; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Grainger & Jacobs, 1996; Norris, 2006; Plaut & Booth, 2000). Of the PDP models, the one that has attracted the most attention among reading researchers in the context of lexical decision is the Plaut and Booth (2000) model. The article introducing this model has more than 60 citations (a considerable achievement in a domain in which competition is fierce) and has been discussed favorably and at considerable length in McNamara's (2005) monograph on semantic priming.

Derek Besner, Szymon Wartak, and Serje Robidoux, Department of Psychology, University of Waterloo, Waterloo, Ontario, Canada.

This research was supported by Natural Sciences and Engineering Research Council of Canada Grant A0998 to Derek Besner. We thank David Plaut for assistance with the simulations and Alan Allport, Dave Balota, Jeff Bowers, Max Coltheart, and David Plaut for their reviews of a previous version of the manuscript.

Correspondence concerning this article should be addressed to Derek Besner, Department of Psychology, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada. E-mail: dbesner@uwaterloo.ca

[1] Joordens, Masson, and Besner (1995) reported simulations with a Hopfield net in which they were unable to simulate additive effects of stimulus quality and word frequency. This failure led them to suggest that discrete stages of processing might be necessary (e.g., see Sternberg, 1998).

This PDP model of lexical processing (here, in the context of the lexical decision task) has been described at some length several times (Plaut & Booth, 2000, 2006); we therefore provide only a brief sketch here. The architecture consists of orthographic and semantic units, with an intervening set of hidden units (see Figure 1 from Plaut & Booth, 2000). Processing feeds forward in a cascaded fasion from orthography to the hidden units and on to semantics; the hidden units and the semantic layer are engaged in interactive activation. The activation function of the units in each layer is sigmoidal. The network is exposed to a training phase in which word pairs are presented with differing frequencies, and a back-propagation algorithm adjusts the connection weights until it settles on a set of weights that allows it to differentiate between items. Details of how a lexical decision is actually computed (over semantics) are described in Plaut and Booth's studies (2000, 2006).

Our initial interest centers on Plaut and Booth's (2000) claim that the model simulates additive effects of stimulus quality and word frequency in lexical decision at the same time that it simulates an interaction between stimulus quality and semantic priming in which targets preceded by an unrelated prime are more impaired by low stimulus quality (relative to high stimulus quality) than are targets preceded by a related prime (Borowsky & Besner, 1993; McDonald, 1980). We also consider the joint effects of stimulus quality and word frequency in lexical decision when there is no formal "prime" event, given that this is the prototypical context in which additive effects of these factors on RT have been observed with skilled readers (see Balota & Abrams, 1995; Becker & Killion, 1977; O'Malley, Reynolds, & Besner, 2007; Norris, 1984[2]; Plourde & Besner, 1997; Stanners, Jastrzembski, & Westbook, 1975; Wilding, 1988[2]; Yap & Balota, 2007; note that Plaut & Booth, 2000, did not report any simulations in this context).

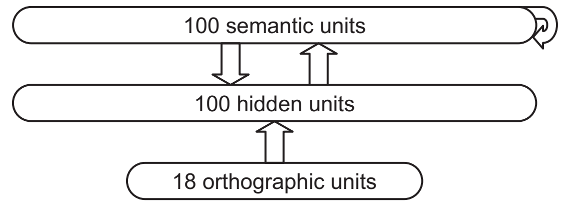

Figure 1. The architecture of the network. Ovals represent groups of units, and arrows represent full connectivity between these groups. Reprinted from "Individual and Developmental Differences in Semantic Priming: Empirical and Computational Support for a Single-Mechanism Account of Lexical Processing," by D. C. Plaut and J. R. Booth, 2000, *Psychological Review, 107,* p. 801. Copyright 2000 by the American Psychological Association.

Table 1
Examples of Additive Effects of Various Factors Reported in the Literature

Lexical decision	Reading aloud
1. Stimulus quality and word frequency Stanners et al. (1975) Becker & Killion (1977) Wilding (1988) Borowsky & Besner (1993) Plourde & Besner (1997) Yap & Balota (2006)	1. Stimulus quality and neighborhood density (for nonwords) Reynolds & Besner (2004)
2. Stimulus quality and semantic priming (when relatedness proportion is .25) Stolz & Neely (1995) Brown, Stolz, & Besner (2006)	2. Stimulus quality and letter length (for nonwords) Besner & Roberts (2003)
3. Stimulus quality and semantic priming (when [spatial] cueing proportion is low) Stolz & Stevanovski (2004)	3. SOA and letter length/whammies (for nonwords) in the PRP paradigm Reynolds & Besner (2006)
4. Semantic priming and [spatial] cueing (when cue proportion is low) Stolz & McCann (2000) Stolz & Stevanovski (2004)	4. SOA and word frequency (in the PRP paradigm) McCann, Remington, & Van Selst (2000)
5. Spatial cueing and word frequency/lexicality McCann, Folk, & Johnston (1992)	
6. SOA and word frequency (in the Psychological Refractory Period [PRP] paradigm) McCann, Remington, & Van Selst (2000)	

Note. From "Visual Language Processing and Additive Effects of Multiple Factors on Timed Performance: A Challenge for the Interactive Activation Framework?" by D. Besner, September 7, 2006, *PsyCrit,* p. 2. Copyright 2006 by Derek Besner. Reprinted with permission. SOA = stimulus onset asynchrony.

On Additivity Between the Joint Effects of Word Frequency and Stimulus Quality in the Plaut and Booth (2000) Model

Plaut and Booth (2000, 2006) simulated a reduction in stimulus quality by reducing the strength of external input to the orthographic units of the model (see Borowsky & Besner's, 2006, p. 183 comments in this regard). Word frequency is manipulated during the training stage of the network only. On each trial, high-frequency words were twice as likely to appear as low-frequency words so that on average the network had twice as much experience with high- as opposed to low-frequency words.

Plaut and Booth (2000) argued that the model simulates additivity between these two factors because of the architecture and particularly because of the processing dynamics associated with the sigmoidal activation function (see Figure 2 from Plaut and Booth, 2000).

Borowsky and Besner (2006) argued that it should therefore not be possible to produce both additive and interactive effects in the same region of the sigmoidal activation function. Additive effects of two factors should be seen if all four points in a 2 × 2 factorial design fall on the linear portion of the activation function, whereas if one or more points are not equidistant from the center relative to the other points, then an interaction should be seen.

[2] One of Wilding's (1988) experiments, with a long intertrial interval, produced an interaction, as did the long intertrial interval condition in Norris's (1984) study. Wilding attributed the interaction to the joint effects of the long foreperiod and attention, arguing that it had nothing to do with reading-related processes.

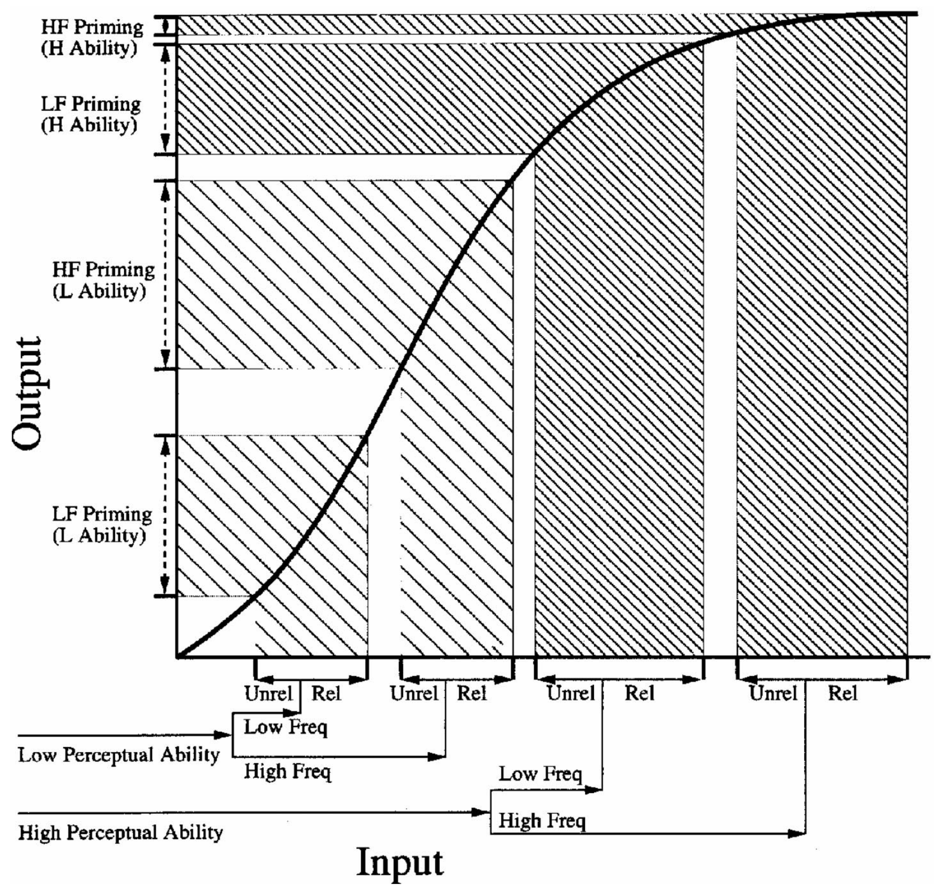

Figure 2. A depiction of how nonlinearities in the sigmoid activation function for semantic units in a distributed attractor network can give rise to greater priming (i.e., the difference in performance following related [Rel] vs. unrelated [Unrel] primes) of low- versus high-frequency target words (LF and HF, respectively) for participants with high perceptual ability (H Ability; narrow-lined regions) but approximately equal priming for low- and high-frequency words for participants with low perceptual ability (L Ability; wide-lined regions). The combination of arrows at the bottom depicts the separate contributions of perceptual ability, target frequency (Freq), and priming context, which are summed together to form the input to a given semantic unit (indicated by the small vertical lines on the *x*-axis), to which the sigmoid function is applied to determine the activation of the unit. Note that relative magnitudes of these contributions are assumed to be greater for high- compared with low-ability participants, greater for high- compared with low-frequency targets, and positive for related primes but negative for unrelated primes (reflecting both facilitation and inhibition, respectively). Moreover, the magnitudes of the contributions of target frequency and priming context are assumed to be greater for high-ability participants because they can process both primes and targets more effectively than low-ability participants. The bottom portion of the sigmoid function is omitted for clarity. Reprinted from "Individual and Developmental Differences in Semantic Priming: Empirical and Computational Support for a Single-Mechanism Account of Lexical Processing," by D. C. Plaut and J. R. Booth, 2000, *Psychological Review, 107,* p. 790. Copyright 2000 by the American Psychological Association.

In reply, Plaut and Booth (2006) reported a simulation in which both additive and interacting effects of several factors are observed when the joint effects of two factors that produce additive effects fall within the range in which one of these factors and a third factor produce an interaction. This apparent success nonetheless led Plaut and Booth (2006) to conclude that

> 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. (p. 199)

In short, this simulation enterprise is running ahead of researchers' current ability to understand what the model is doing in detail. Perhaps this is an inevitable stage in the development of computational models with distributed representations. It is nonetheless disquieting not to be able to understand such details.

Replication

The first order of business here is to show that, in our hands, the Plaut and Booth (2000) model replicates the data reported by Plaut and Booth (2006). We therefore started by running the model with Plaut and Booth's (2006) stimulus onset asynchrony (SOA) and stimulus duration parameters in order to examine the joint effects of stimulus quality and word frequency, and stimulus quality and semantic relatedness. The top panel of Table 2 shows the data from Plaut and Booth (2006). The middle panel depicts the data obtained when we ran the model. The bottom panel shows the difference scores between their data and the present results. The largest difference between the Plaut and Booth (2000) results and the present simulation is .002 units (RTs from the model). The results of the present simulations therefore appear to be a good (if not perfect) match to those reported by Plaut and Booth (2006).[3]

These data are also reported in Figure 3 in a different format so as to make it easier for the reader to appreciate the pattern. Given that the magnitude of some of the effects is very small, the magnitude of the interaction (represented by a single point) between stimulus quality and word frequency is plotted in the left-hand panel, and the magnitude of the interaction between stimulus quality and semantic relatedness is plotted in the right-hand panel. Given this successful replication, we now consider Plaut and Booth's (2000, 2006) claim that the model produces additive effects of stimulus quality and word frequency.

An Underadditive Trend Between Stimulus Quality and Word Frequency

Our first comment concerns the unusual trend Besner and Borowsky (2006) noted in the Plaut and Booth (2006) simulation data and that we reproduced here in the left-hand panel of Figure 3. That is, low stimulus quality slows high-frequency words more than do low-frequency words (relative to the high stimulus quality condition)—a pattern that has never been reported in studies of skilled readers. This underadditivity is not significant (despite being nearly twice the magnitude of the significant interaction between semantic relatedness and stimulus quality seen in the right-hand panel). Nonetheless, we take this as a sign that the model may not capture what it is that skilled readers produce. We return to this issue when we consider the joint effects of stimulus quality and word frequency in the absence of a prime event.

Table 2
Settling Times (in Cycles) for Target Words in the Model as Reported by Plaut and Booth (2006) and in the Present Simulations

Activation strength	High frequency		Low frequency	
	Related	Unrelated	Related	Unrelated
Plaut and Booth (2006)				
0.90	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
Present simulations				
0.90	4.436	4.482	4.609	4.683
0.82	4.463	4.516	4.625	4.705
0.75	4.496	4.554	4.658	4.747
Difference scores				
0.90	−0.001	−0.001	0.000	−0.001
0.82	0.000	0.000	0.000	0.000
0.75	−0.001	−0.001	0.000	−0.002

How Is the Size of the Stimulus Quality Effect in the Model Related to the Size of the Stimulus Quality Effect in Skilled Readers?

It is also apparent from Figure 3 that virtually perfect additivity is observed when reduced activation strength (representing lowered stimulus quality) is .75. This is the largest reduction in stimulus quality that Plaut and Booth (2006) tested in response to Borowsky and Besner's (2006) concern that the range of stimulus quality tested by Plaut and Booth (2000) was very small indeed (.82 [reduced activation] vs. full activation of .90).

We remain unsatisfied by the small range tested to date, given that stimulus quality and word frequency have additive effects in skilled readers even with remarkably large stimulus quality manipulations. For example, Stanners et al. (1975), Yap and Balota (2007), and O'Malley et al. (2007), each of whom used a different method of reducing stimulus quality, all reported additive effects of stimulus quality and word frequency when the main effect of stimulus quality is at least 100 ms. Indeed, Borowsky and Besner (1993) reported statistically additive effects of stimulus quality and word frequency when the main effect of stimulus quality was over 200 ms.

Plaut and Booth's (2006) Estimates of the Magnitude of the Stimulus Quality Effect

In their initial report, Plaut and Booth (2000, p. 803) conducted a simple linear regression to relate their model's settling times to human RTs. Using the results of this regression, they arrived at an estimate of 35.3 ms for the magnitude of the stimulus quality effect in the model when 0.82 was the activation strength on low stimulus quality trials. In response to Borowsky and Besner's (2006) concern about the small size of the stimulus quality manipulation, Plaut and Booth (2006) replicated the simulation using 0.82 as the input strength and added a simulation using a stronger stimulus quality manipulation (reducing the input strength in the low stimulus quality condition from 0.82 to 0.75). Observing that reducing the input strength to 0.75 resulted in a 2.54-fold increase in the magnitude of the stimulus quality effect in model RTs, Plaut and Booth (2006) multiplied their original estimate of 35.3 (based on the simulations in 2000) by 2.54 (based on the simulations in 2006) to arrive at a new estimate of 89.7 ms in terms of human RTs.

[3] The reader may wonder why any discrepancies exist at all. Though we are unable to verify 100% of the output from our simulations, Plaut and Booth (2006) kindly provided a sample of their output from the simulations conducted in 2006. For that sample, the output from our replication simulations matched exactly. Given this consistency in the two sets of simulation outputs, we assume that the discrepancies arise from differences in the details of the postsimulation analyses—we did our best to match the procedures described in Plaut and Booth's (2000, 2006) studies, but slight variations may have been introduced where the procedures were less clearly specified.

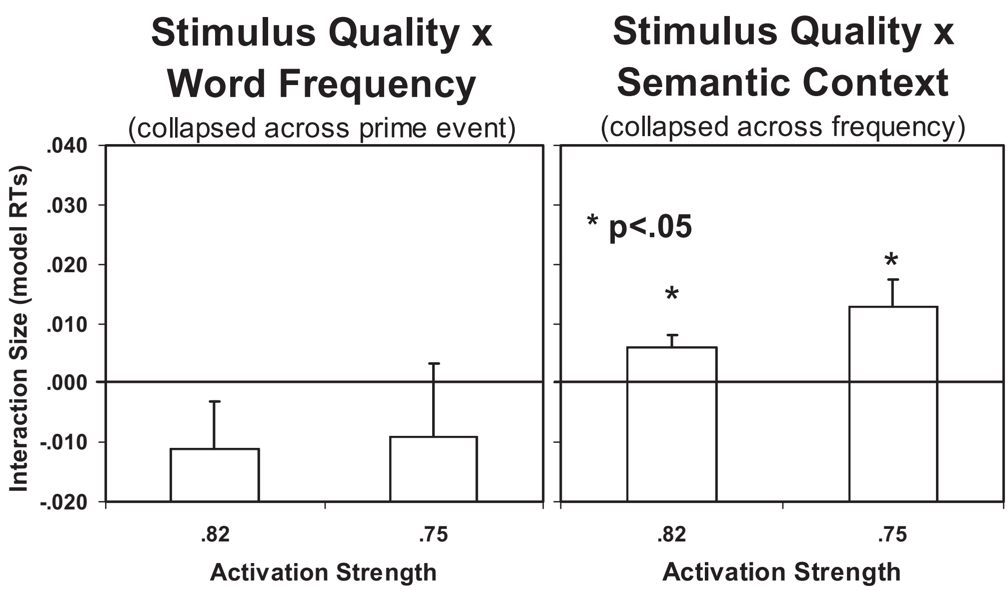

Figure 3. Replication of Plaut and Booth's (2006) simulations. Following Plaut and Booth (2006), these data are collapsed across both prime-target stimulus onset asynchrony and prime duration. The interaction scores between stimulus quality and word frequency can be seen in the left panel (collapsed across semantic context) and between stimulus quality and semantic context in the right panel (collapsed across word frequency). The horizontal line at 0 represents perfect additivity. Values below the line represent underadditivity, and values above the line represent overadditivity. The standard error associated with each interaction is indicated by the error bar.

Given that the data from the simulations in Plaut and Booth's (2000) study were not reported and have since been lost (Plaut & Booth, 2006, footnote p. 197), it would seem prudent to redo the regressions to ensure that the new simulations conducted in 2006 replicate the estimates from the original simulations for the 0.82 versus the 0.90 simulations before using the new simulation (0.75) to scale the estimate.

Our Estimates of the Magnitude of the Stimulus Quality Effect

Here we undertake to provide estimates of the magnitude of the stimulus quality effect at various levels of input strength based only on the current simulations. One approach might be to treat each level of low input strength as a separate experiment and thus use those data to produce estimates of the magnitude of the stimulus quality manipulation separately for each experiment. However, because this approach yields regression equations that differ considerably from each other, the implication would be that the relationship between model time and human RT differs depending on the level of low input strength.[4] A better approach is to choose one condition of low input strength (low stimulus quality) and apply the resulting regression formula to all other simulations. There is no theoretical reason to choose one level of input strength over another as best representing the human experiment conducted in Plaut and Booth's (2000) study. Consequently, we conducted individual regression analyses for each level of low input strength and regressed the model's reported RTs against the adult data reported in Plaut and Booth's (2000) study.[5] In all cases, the same simulation data (that for an input strength of 0.90) was used to reflect the "high perceptual ability" data. We then used each of these regression estimates to produce estimates of the effect at all levels of input strength. The result was four estimates of the magnitude of the stimulus quality manipulation for each level of input strength simulated. (In the interest of completeness, we conducted the regression analyses for both levels of input strength reported in Plaut and Booth's, 2006, study as well as for the additional simulations presented later in this article.) Because the model under discussion is meant to represent a skilled reader, we used the human RTs reported in Plaut and Booth's (2000) experiments with adult participants. Table 3 provides a summary of the regression analyses when fitting models of the form $Y = mX + b$, where Y is the human RT data from each condition reported in Plaut and Booth's (2000, Appendix B) study, X is the corresponding model settling time for each condition, m is the slope, and b is the intercept.

First, it is important to note that the 35.3-ms stimulus quality effect in the 0.82 condition that was reported in Plaut and Booth's (2000) study was not replicated using the simulation

[4] We thank Dave Plaut for pointing out this problem with the described approach.

[5] Because the results we are discussing are taken from Plaut and Booth's (2006) study, in which the analysis was collapsed across SOA, the human data used here represent the collapsing of the adult data from Experiments 1 (long SOA) and 3 (short SOA) in Plaut and Booth's (2000) study.

data from 2006,[6] regardless of which model data are used to represent the low stimulus quality condition. Second, Plaut and Booth (2000) used the human data and simulation data from their Experiment 1 and applied the resulting coefficients to all subsequent estimates. This seems a somewhat arbitrary choice (Why Experiment 1 rather than Experiment 3?). The present analysis presents estimates of the stimulus quality effect size for all conditions, using each simulation as the baseline or Plaut and Booth's (2000) equivalent condition in turn. The key point is that regardless of which simulation is used to represent the human results in Plaut and Booth's (2000) study, the estimates of the stimulus quality manipulations remain well within the range of stimulus quality effects reported in the literature. Indeed, the largest estimate of the stimulus quality effect size is 109 ms, which is very close to where Stanners et al. (1975), O'Malley et al. (2007), and Yap and Balota (2007) all reported statistically additive effects of stimulus quality and word frequency.

These new regression results also highlight another problem with these estimates. That is, a theoretical dilemma is raised by the large negative intercept coefficients reported in Table 3 (where low stimulus quality is represented by activation strengths of 0.82, 0.75, 0.70). The intercept can be thought of as the amount of time that would be required to produce a response in the absence of any of the processing presumably described by the model. Under this interpretation, the intercept should encapsulate any precognitive and postcognitive processing or motor-response times. If one ignores extrasensory perception for the moment, it is theoretically untenable that these values would be negative. It is perhaps more reasonable to constrain the intercept to be no smaller than 0 and use those slopes to estimate human RTs. Fitting linear regressions to models of the form $Y = mX + 0$ (i.e., constraining the intercept to be 0) produces the results in Table 4. The estimates of the stimulus quality effects are now reduced in these models (and are remarkably stable).[7]

Given this regression analysis, we are confident that decreasing the input strength to as low as 0.65 provides a better test of the model's performance across the range of stimulus quality manipulations found in the literature with skilled readers.

Table 3
Results From Regressing Human Reaction Times Against Model Settling Times and Estimates of the Magnitude of the Stimulus Quality (SQ) Manipulation

Alternative analysis	Slope (*m*)	Intercept (*b*)	Estimated SQ effect sizes			
			0.82	0.75	0.70	0.65
Raw model settling times[a]			0.025	0.061	0.154	0.385
Model data used for regression[b]						
Low input = 0.82	230	−348	6 ms	14 ms	35 ms	89 ms
Low input = 0.75	284	−598	7 ms	17 ms	44 ms	109 ms
Low input = 0.70	246	−438	6 ms	15 ms	38 ms	95 ms
Low input = 0.65	139	42	3 ms	8 ms	21 ms	54 ms

Note. Models of the form $Y = mX + b$, where Y is the human reaction time data from each condition reported in Plaut and Booth's (2000, Appendix B) study, and X is the corresponding model settling time for each condition.
[a] The SQ effect sizes for the raw model settling times are the difference in settling times between the high and low activation strength conditions, collapsed across semantic context and word frequency. Human RTs are estimated from the model parameters (slope and intercept). [b] All models used 0.90 as the baseline (high stimulus quality).

Table 4
Results From Regressing Human Reaction Times Against Model Settling Times and Estimates of the Magnitude of the Stimulus Quality (SQ) Manipulation

Alternative analysis	Slope (*m*)	Intercept (*b*)	Estimated SQ effect sizes			
			0.82	0.75	0.70	0.65
Raw model settling times[a]			0.025	0.061	0.154	0.385
Model data used for regression[b]						
Low input = 0.82	154	0	4 ms	9 ms	24 ms	59 ms
Low input = 0.75	153	0	4 ms	9 ms	24 ms	59 ms
Low input = 0.70	152	0	4 ms	9 ms	23 ms	59 ms
Low input = 0.65	148	0	4 ms	9 ms	23 ms	57 ms

Note. Models of the form $Y = mX$ (intercept constrained to 0), where Y is the human reaction time data from each condition reported in Plaut and Booth's (2000, Appendix B) study, and X is the corresponding model settling time for each condition.
[a] The SQ effect sizes for the raw model settling times are the difference in settling times between the high and low activation strength conditions, collapsed across semantic context and word frequency. Human RTs are estimated from the model parameters (slope and intercept). [b] All models used 0.90 as the baseline (high stimulus quality).

The Joint Effects of Stimulus Quality and Word Frequency in the Presence of a Prime Event

The results of this simulation (collapsed across SOA, prime duration, and semantic relatedness) can be seen on the right-hand side of Figure 4. We confirm Plaut and Booth's (2006) report that at 0.82 and 0.75 the interaction between stimulus quality and word frequency is not significant (0.82: $F[1, 126] = 1.88$, $p > .10$, $MSE = .002$; 0.75: $F[1, 126] < 1$, $p > .40$, $MSE = .005$). However, when the strength of the external input to the model is reduced beyond .75 (0.70 and 0.65), the model now produces a significant interaction between stimulus quality and word frequency that has typically not been observed in the literature on

[6] Our interpretation of the procedures described in Plaut and Booth's (2000) study suggests that the human adult data collapsed across Experiments 1 and 3 are most appropriate to the present analysis. However, to reassure ourselves that the failure to produce results more similar to theirs was not simply a matter of using the wrong human data, we conducted similar regressions with the children's RTs and with the adult RTs on the subset of items presented to the children. When we used the children's data, we fared only marginally better, estimating the stimulus quality effect at approximately 9 ms.

[7] Constraining the models in this way reduces the adjusted r^2. This is to be expected because models of the form $Y = mX + b$ have two freely varying parameters (m and b), whereas models of the form $Y = mX$ have only one (m). However, the benefit of having the extra parameter must be weighed against the cost of having a model that is theoretically impossible.

skilled readers (0.70: $F[1, 125] = 4.09$, $p < .05$, $MSE = .069$; 0.65: $F[1, 124] = 12.55$, $p < .005$, $MSE = .280$).

The Joint Effect of Stimulus Quality and Word Frequency When There Is No Prime Event

The joint effects of stimulus quality and word frequency have most often been examined in the context of a so-called "running" lexical decision, in which there is no prime event between targets (e.g., Balota & Abrams, 1995; Becker & Killion, 1977; Norris, 1984; Plourde & Besner, 1997; O'Malley et al., 2007; Stanners et al., 1975; Wilding, 1988; Yap & Balota, 2007). In our final set of simulations, we considered the effect of the full range of stimulus quality tested here (.90, .82, .75, .70, .65) on performance when there is no prime event. The absence of a prime event in the model eliminates the variance in target processing attributable to prime processing, which should increase power to detect the underadditive interaction trend noted earlier when the reduction in the size of the stimulus quality effect is small. These data can be seen in Figure 5.

When the reduction in stimulus quality is small (0.82 and 0.75 input strength; the levels reported in Plaut and Booth's, 2006, study), high-frequency words are *more* impaired than low-frequency words (0.82: $F[1, 123] = 6.21$, $p < .05$, $MSE = .002$; 0.75: $F[1, 122] = 4.32$, $p < .05$, $MSE = .004$). We emphasize, again, that this underadditive pattern has never been reported in studies of skilled readers.

When the reduction in stimulus quality is slightly larger (.70 input strength), additivity is seen, $F(1, 122) = 1.37$, $p > .20$, $MSE = .044$. However, as the reduction in stimulus quality increases (.65 input strength), low-frequency words are more impaired than high-frequency words, $F(1, 118) = 8.22$, $p < .01$, $MSE = .159$).[8]

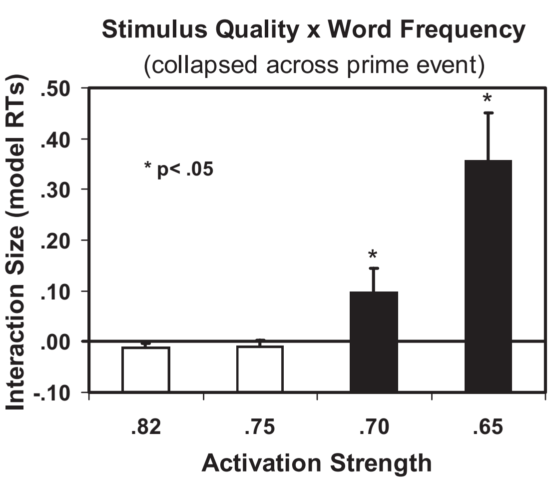

Figure 4. Following Plaut and Booth (2006), these data are collapsed across both prime-target stimulus onset asynchrony and prime duration. Depicted here are interaction scores between stimulus quality and word frequency, with the range of activation strength extended to include .70 and .65. The horizontal line at 0 represents perfect additivity. Values below the line represent underadditivity, and values above the line represent overadditivity. The white bars are reproduced from the left panel of Figure 3 (note that they are on a different scale), and the black bars represent stronger stimulus quality manipulations. The standard error associated with each interaction is indicated by the error bar.

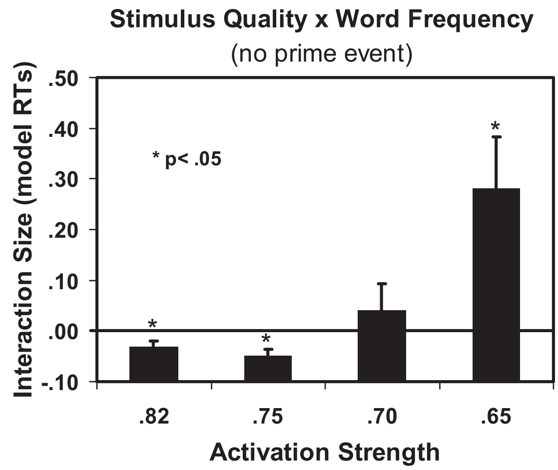

Figure 5. The interaction scores between stimulus quality and word frequency when there is no prime event. The horizontal line at 0 represents perfect additivity. Values below the line represent underadditivity, and values above the line represent overadditivity. The standard error associated with each interaction is indicated by the error bar. RT = reaction time.

In short, the joint effects of stimulus quality and word frequency in the Plaut and Booth (2000) model are not constant. Instead, three patterns of performance (underadditive, additive, and overadditive) are observed as the size of the stimulus quality manipulation increases from small to large. The literature on skilled readers displays no such trend. Instead, across a wide range, the joint effects of stimulus quality and word frequency are statistically additive.

General Discussion

It is not immediately obvious why this implemented model fails to produce systematic additivity. One possibility is that the presence of feedback between the hidden units and semantics works against it. If this is true then interactive activation (IA), currently seen as a fundamental principle of cognition in some quarters, would be rather more limited (indeed, a number of researchers argue that IA is more limited than generally acknowledged: e.g., see Borowsky & Besner, 1993; Brown, Stolz, & Besner, 2006; O'Malley et al., 2007; Robidoux, Stolz, & Besner, 2007; Stolz & Neely, 1995; Smith & Besner, 2001; Yap & Balota, 2007). Further simulation work that addresses the general issue of whether it is possible to obtain additive effects of two factors in the presence of processes engaged in interactive activation would therefore appear to be an important goal.

Another possibility is that the ratio of cascade rates across different levels needs consideration given what is known about the

[8] When we applied the outlier removal procedure described in Plaut and Booth's (2000, 2006) studies to the output from the "no prime" simulations, some items were left with no trials remaining in certain cells. These items were not included in the analysis, resulting in the decreased degrees of freedom reported here.

conditions that produce additivity in the context of models that are feedforward only (see McClelland, 1979; Roberts & Sternberg, 1993). Perhaps cascade rates and IA play a joint role in making systematic additivity so difficult to obtain. Whatever the answer(s), it seems clear that there are qualitative differences between the implemented model and the data from skilled readers that undermine the Plaut and Booth (2000) model in its current form.

It should not go unmentioned that other investigators (e.g., Borowsky & Besner, 1993; O'Malley et al., 2007; Plourde & Besner, 1997; Yap & Balota, 2007) have suggested a rather different approach to explaining why stimulus quality and word frequency are additive in the context of lexical decision. They proposed that the output from a processing stage affected by stimulus quality is thresholded and that the effect(s) of word frequency occur further downstream. This approach represents a radical departure from the assumptions underlying all computational models currently on the table (e.g., Coltheart et al., 2001; Harm & Seidenberg, 2004; Perry et al., 2007; Plaut et al., 1996; Seidenberg & McClelland, 1989), but of course it is based on Sternberg's seminal idea of serially organized and discrete processing stages (Roberts & Sternberg, 1993; Sternberg, 1969, 1998, 2001). To be sure, the investigators proposing thresholded processing so as to account for additive effects of stimulus quality and word frequency in the context of lexical decision (and in particular, with nonword foils that are orthographically legal and that are not pseudohomophones) restrict this assumption to early processing. It remains to be seen whether other processes are best explained in terms of a cascade, IA, or thresholding.

Is Additivity of Stimulus Quality and Word Frequency Restricted to the Lexical Decision Task?

Yap and Balota (2007) and O'Malley et al. (2007) have reported that this pattern of additivity between stimulus quality and word frequency is not observed in other tasks; in both reading aloud and semantic categorization, these two factors interact. This observation might incline some theorists to view the lexical decision data (and task?) as less important than, for example, the reading aloud data. It turns out, however, that the interaction observed in reading aloud is not task specific but is instead modulated by the presence versus absence of nonwords in the stimulus list. Additivity is again observed in reading aloud when nonwords are mixed with the words (O'Malley & Besner, 2007).

Other Examples of Additive Effects

We have restricted the simulation work in the present article to the joint effects of stimulus quality and word frequency in the context of lexical decision. However, there are other examples of additive effects in the visual word recognition literature that are not restricted to this task (see Table 1 from Besner, 2006).

Beyond the examples summarized by Besner (2006), there are a number of reports that the interaction between stimulus quality and semantic priming (or relatedness) is modulated by strength of association[9] and by relatedness proportion[10] in both lexical decision (Brown et al., 2006; Robidoux et al., 2007; Stolz & Neely, 1995) and reading aloud (Ferguson, Robidoux, & Besner, 2007). When strength of association and/or relatedness proportion is low, semantic relatedness and stimulus quality are additive factors in both tasks. Such results are likely problematic for both the Plaut and Booth (2000) model and other IA computational models currently on the table, given that there are few restrictions to IA in the latter cases and our suspicion that IA models have difficulty producing additivity.

Conclusion

There is currently no evidence that any computational model of visual word recognition and reading aloud on the table is able to simulate systematically additive effects of two factors on RT. This situation exists despite the fact that the first instance of additive effects in this domain was reported over a quarter of a century ago.

Coltheart et al. (2001) and Perry et al. (2007) have argued that in order to assess the success of a computational model, researchers need to develop a set of benchmarks against which performance can be measured. We suggest that many of the examples of additivity reported in the literature are overdue to be considered benchmarks.

[9] The interaction is present when strength of association between prime and target is relatively strong, but it disappears when the strength of association is weak, regardless of whether the strength of association manipulation (strong vs. weak) is blocked (Stolz & Neely, 1995) or intermixed (Robidoux et al., 2007).

[10] In both of the cited experiments, the interaction between relatedness and stimulus quality is eliminated when only 25% of word trials are related.

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., & 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. (2006). *Visual language processing and additive effects of multiple factors on timed performance: A challenge for the interactive activation framework?* Retrieved July 16, 2007, from *PsyCrit* Web site: http://psycrit.com/wikiup/6/6a/Besner2006.pdf

Besner, D., & Borowsky, R. (2006). Postscript: Plaut and Booth's (2006) new simulations: What have we learned? *Psychological Review, 113,* 194–195.

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. (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,* 181–194.

Brown, M., Stolz, J. A., & Besner, D. (2006). Dissociative effects of stimulus quality on semantic and morphological context effects in visual word recognition. *Canadian Journal of Experimental Psychology, 60,* 190–199.

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

Ferguson, R., Robidoux, S., & Besner, D. (2007). *Reading aloud: On the joint effects of semantic context and stimulus quality as modulated by relatedness proportion*. Manuscript in progress.

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

Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review, 111,* 662–720.

Joordens, S., Masson, M. E. J., & Besner, D. (1995, November). *Connectionist models and additive effects: Are distinct stages of processing necessary?* Poster session presented at the annual meeting of the Psychonomic Society, Los Angeles.

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of 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. 3–36). Hillsdale, NJ, England: Erlbaum.

McClelland, J. L. (1991). Stochastic interactive processes and the effect of context on perception. *Cognitive Psychology, 23,* 1–44.

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

McDonald, J. E. (1980). *An information processing analysis of word recognition.* Unpublished doctoral dissertation, New Mexico State University.

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

Norris, D. (1984). The effects of frequency, repetition and stimulus quality in visual word recognition. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 36*(A), 507–518.

Norris, D. (2006). The Bayesian reader: Explaining word recognition as an optimal Bayesian decision process. *Psychological Rreview, 113,* 327–357.

O'Malley, S., & Besner, D. (2007, June). *Visual word recognition: Are the processing dynamics fixed?* Paper presented at the annual meeting of the Canadian Society for Brain, Behaviour, and Cognitive Science, Victoria, British Columbia, Canada.

O'Malley, S., Reynolds, M. G., & Besner, D. (2007). Qualitative differences between the joint effects of stimulus quality and word frequency in reading aloud and lexical decision: Extensions to Yap and Balota (2007). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 451–458.

Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: The CDP+ model of reading aloud. *Psychological Review, 114,* 273–315.

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.

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

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

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.

Robidoux, S., Stolz, J. A., & Besner, D. (2007, June). *Visual word recognition: Control over interactive activation*. Paper presented at the annual meeting of the Canadian Society for Brain, Behaviour, and Cognitive Sciences, Victoria, British Columbia, Canada.

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach.* Cambridge, MA: MIT Press.

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

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

Stanners, R. F., Jastrzembski, 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. *Acta Psychologica, Amsterdam, 30,* 276–315.

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: MIT Press.

Sternberg, S. (2001). Separate modifiability, mental modules, and the use of pure and composite measures to reveal them. *Acta Psychologica, 106,* 147–246.

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

Wilding, J. M. (1988). The interaction of word frequency and stimulus quality in the lexical decision task: Now you see it, now you don't. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 40*(A), 757–770.

Yap, M. J., & Balota, D. A. (2007). Additive and interactive effects on response time distributions in visual word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 274–296.

Received March 19, 2007
Revision received July 13, 2007
Accepted July 16, 2007 ■