SD-Squared Revisited: Reply to Coltheart, Tree, and Saunders (2010)

Anna M. Woollams and Matthew A. Lambon Ralph
University of Manchester

David C. Plaut
Carnegie Mellon University

Karalyn Patterson
Medical Research Council Cognition and Brain Sciences Unit

The connectionist triangle model of reading aloud proposes that semantic activation of phonology is particularly important for correct pronunciation of low-frequency exception words. Our consideration of this issue (Woollams, Lambon Ralph, Plaut, & Patterson, 2007) reported computational simulations demonstrating that reduction and disruption of this semantic activation resulted in the marked deficit in low-frequency exception word reading that is characteristic of surface dyslexia. We then presented 100 observations of reading aloud from 51 patients with *semantic dementia* (SD) demonstrating a universal decline into *surface dyslexia,* a phenomenon we termed "SD-squared." Coltheart, Tree, and Saunders (2010) have more recently provided a simulation of the SD-squared data within the dual route cascaded (DRC) model, achieved by varying the amount of damage to components of the lexical and nonlexical pathways. Although they suggested that these simulations provide a closer fit to the SD patients' reading data than our own, we demonstrate here that this is not the case. Moreover, we argue that the connectionist triangle model account has substantially greater explanatory and predictive power than the DRC account.

Keywords: surface dyslexia, semantic memory, reading aloud, frequency, regularity

The role of semantic information in normal reading aloud is a matter of considerable debate and a dimension along which current computational models vary widely. In connectionist triangle models (e.g., Plaut, McClelland, Seidenberg, & Patterson, 1996), which eschew a structural level of lexical nonsemantic representation, the semantic pathway (orthography to semantics to phonology: O→S→P) is the primary source of whole-word knowledge in reading aloud and can supplement activation from the direct pathway (O→P) to ensure correct production of words with exceptional spelling–sound correspondences. This view has been supported by multiple reports of a strong association between level of semantic knowledge and exception word reading amongst patients with semantic dementia (SD) who are characterized by progressive deterioration of meaning-level representations (Fushimi, Komori,

Ikeda, Lambon Ralph, & Patterson, 2009; Graham, Patterson, & Hodges, 2000; Jefferies, Lambon Ralph, Jones, Bateman, & Patterson, 2004; Patterson & Hodges, 1992; Patterson et al., 2006; Woollams, Lambon Ralph, Plaut, & Patterson, 2007). More recently, the extent of the anterior temporal lobe atrophy that causes SD has also been shown to correlate specifically with accuracy of exception word reading (Brambati, Ogar, Nehaus, Miller, & Gorno-Tempini, 2009).

A contrasting view is provided by localist models of reading aloud, such as the dual route cascaded (DRC) model, which incorporate one or more structural levels of lexical nonsemantic representation (e.g., Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001), obviating the need for involvement of semantic information to support the correct pronunciation of exception words. Advocates of this approach cite some (relatively rare) observations of classical single dissociations between semantic knowledge and exception word reading ability as apparent evidence for the separation of lexical and semantic knowledge (e.g., Coltheart, 2006). Within this account, the strong association between level of semantic knowledge and exception word reading seen in SD does not imply a causal relationship but rather is mediated by a third factor, which is the spread of atrophy from the anterior temporal regions involved in semantic processing to other largely unspecified neural regions that support correct exception word reading (Coltheart, Tree, & Saunders, 2010).

We (Woollams et al., 2007) provided a simulation of the expected consequences of the semantic degradation observed in SD for reading aloud within the connectionist triangle framework of Plaut et al. (1996). Processing of low-frequency and/or exception words within this model comes to rely in part upon semantic activation of phonology ("S"→P) provided during training, pro-

ducing a graded division of labor between the direct and semantic pathways. We designate semantics as "S" instead of S to reflect the fact that the Woollams et al. (2007) simulations did not include an implementation of semantics but only a frequency-weighted approximation of its contribution to phonology. This approach was meant to capture the theoretical commitment of the connectionist triangle account that whole-word activation of phonology, when it occurs, derives primarily from semantics. In line with this view, a comparable graded division of labor is also apparent in connectionist models that have included a fully implemented semantic system (Harm & Seidenberg, 2004).

In our lesion simulation (Woollams et al., 2007), we both weakened "S"→P activation and added a constant level of inverse frequency-weighted noise to this signal, the impact of which became more apparent with increasing semantic damage. The addition of noise was motivated by the observation that the SD patients' picture-naming errors are not always omissions but also include errors of commission (Hodges, Graham, & Patterson, 1995; Lambon Ralph, Graham, Ellis, & Hodges, 1998; Rogers et al., 2004; Woollams, Cooper-Pye, Hodges, & Patterson, 2008). The inevitable consequence of compromising "S"→P activation in our lesion simulation was a deficit in the reading of low-frequency exception words that defines surface dyslexia. In addition, we modeled occasional classical single dissociations between level of semantic knowledge and exception word reading accuracy as extremes in the normal distribution of the network's premorbid division of labor. This meant that networks provided with weak "S"→P activation during training would require a relatively large reduction in this source of activation before exception word reading began to suffer, and conversely, in networks provided with strong "S"→P activation during training, a relatively small reduction would impair performance on these items.

We then went on to assess the adequacy of our predictions with reference to the largest ever case-series consideration of reading aloud in SD, involving 100 observations from 51 patients. These data clearly demonstrated the overwhelming association between level of semantic knowledge and low-frequency exception word reading, with half of the variance in the latter accounted for by a composite semantic score representing the simple average of picture naming and spoken word to picture matching, neither of which involves any reading whatsoever. We did nevertheless observe a few dissociations between level of semantic knowledge and low-frequency exception word reading accuracy: Three cases with an appreciable semantic impairment were initially unimpaired in their reading, and two cases with an as-yet relatively mild semantic impairment manifested a reading deficit. The low incidence of such dissociations (5/100) was consistent with their being outliers within a normal distribution of premorbid division of labor, as postulated in our simulations. Longitudinal semantic and reading data were available for these five cases, and with increasing semantic impairment, low-frequency exception word reading for all of these cases declined, resulting in a universal pattern of surface dyslexic reading. This progression is entirely in accordance with our simulations, demonstrating that what varies across cases is simply the degree of semantic deterioration required before surface dyslexia emerges: This occurs later for individuals whose premorbid reading relied less on additional phonological activation from whole-word semantic knowledge and earlier amongst those whose reading relied more on this source of information.

We suggested that our triangle model simulation provided a principled and parsimonious account not only of the central tendency observed in the patient data in terms of a general division of labor between the direct and semantic pathways of the reading system but also of the variability around this central tendency in terms of premorbid variation in the degree of this division of labor. Nevertheless, we explicitly acknowledged (Woollams et al., 2007, p. 333) that it should be possible to simulate the patient data we reported in alternative frameworks such as the DRC model by assuming multiple impairments to the (unimplemented) semantic system, the lexical route, and the nonlexical route to explain the central tendency observed. The explanation of outlier observations within this model is via relative preservation of one or more of these components.

In their comment, Coltheart et al. (2010) have simulated the SD-squared pattern within the DRC model following the general method we suggested in our original article. In short, they used variations in the extent of damage to the functionally independent direct lexical and nonlexical pathways of the model to generate a very large (40,200) set of models. From these, they selected the particular combination of lexical and nonlexical damage that most closely matched each of the 100 observations of reading aloud data that we reported. A regression of the reading data from these models on the SD patients' composite semantic score produced a pattern similar to that seen when the patients' own reading scores were regressed onto their composite semantic score. Coltheart et al. argued that their DRC simulations provide a better fit to the SD patient data than our own triangle model simulations and hence that the DRC model of reading aloud is to be preferred to the triangle model as an adequate explanation of both normal and impaired performance.

In this reply, we demonstrate that the conclusion of Coltheart et al. (2010) is unjustified on two fronts. The first domain of disagreement is factual, in that many of the claims made by Coltheart et al. concerning interpretation of the patients' reading performance, the nature of their underlying neural damage, and the adequacy of the DRC simulation data relative to that of the triangle model are invalid. The second and somewhat independent domain of disagreement is theoretical in nature, as it concerns the criteria that should be used to evaluate the adequacy of the competing accounts provided by the different models. It is our view that the litmus test of theoretical adequacy is not purely a model's ability to yield an exact reproduction of empirical data: In fact, we show that to prioritize this goal above all others can have undesirable consequences. Instead, we believe that any adequate theoretical account of normal and disordered reading aloud must provide an explanation of the data that it simulates in terms of neuroanatomical bases and relationships to other cognitive domains, thereby providing convergent evidence and generating novel empirical predictions. On both factual and theoretical fronts, we argue that the DRC simulations reported by Coltheart et al. do not challenge the superior explanatory capacity of the connectionist triangle account of surface dyslexic reading.

Data Interpretation

The Notion of the Three Phases of Reading Aloud in SD

At the outset of their article, Coltheart et al. (2010) suggested that, on the basis of the data we reported, three phases may be

identified in the reading aloud behavior of any given SD patient: (a) intact, (b) pure surface dyslexia, and (c) generalized impairment. We do not dispute that all three patterns of reading behavior may be observed across a group of SD patients; we do, however, challenge Coltheart et al.'s proposal that each stage emerges in sequence as the disease progresses. This commitment to sequential progression implies a positive relationship between reading phase and severity of the disease. From Coltheart et al.'s description, it is not clear what might be used as an independent variable to classify any given patient into one of these phases, yet one must exist, as to classify only on the basis of reading performance makes the definition of these phases entirely circular. In SD, the best marker for disease severity, and indeed, one that directly correlates with the level of anterior temporal atrophy that defines it, is performance on semantic measures (Mummery et al., 2000), such as the picture-naming and spoken-word-to-picture-matching tests that we used to compute a composite semantic score. Indeed, as we consider further later, Coltheart et al. chose to use this same variable as a measure of severity that they then employed as an independent predictor in their reading simulations.

By this logic then, if we were to divide our patient data into three levels of severity according to composite semantic score, we should observe these three stages of reading aloud behavior emerging in turn. We have done this in Figure 1, where the continuum of semantic impairment has been divided into three levels of severity using two vertical lines. The horizontal line of any given graph marks the two standard deviations below normal control performance that may be used to dichotomize performance as intact or impaired. The symbols used for each observation represent the phase that would be assigned to each within the Coltheart et al. (2010) scheme. Specifically, those cases with universally intact reading aloud are termed Phase 1, those cases with impaired exception word reading but intact regular word reading are termed Phase 2, and those cases with impaired reading of both exception words and regular words are termed Phase 3. This treatment is consistent with Coltheart et al.'s (2010, p. 259) proposal that impairments in regular word reading and nonword reading both arise from damage to the nonlexical route associated with Phase 3.

As is immediately apparent from Figure 1, observations representing each of the different phases of reading aloud in semantic dementia are distributed across at least two, if not all three, levels of severity of the semantic impairment: Cases of intact reading are not limited to the milder range, pure surface dyslexic cases do not all appear within the moderate range, and generalized reading impairments are not confined to the severe range. Put another way, mild patients may be in Phase 1 or Phase 2; moderate patients may be in Phase 1, Phase 2, or Phase 3; and severe patients may be in Phase 2 or Phase 3. The imposition of this artificial trichotomy of phases of SD reading onto what is clearly a continuous distribution with individual variation is therefore a misrepresentation of the patient data that we reported. Within our account, variation in reading impairment at any given level of disease severity is attributed to individual differences in premorbid division of labor. Although this variation was simulated by Coltheart et al. (2010) by applying different amounts lexical and nonlexical damage, these levels of damage were meant to correspond to the severity of the disorder within their account. It is therefore not clear to us on what grounds Coltheart et al. assigned different levels of damage to

components of the DRC model to simulate the performance of patients of equivalent severity.

Insights From Imaging: The Neuroanatomy of Reading Aloud in SD

A great deal of recent neuroimaging research and also many years of cognitive neuropsychology and behavioral neurology are based on the assumption that cognitive abilities like reading aloud are at least partially localized to specific regions of the brain. The implication of Coltheart et al.'s (2010) three stages of reading aloud in SD is that there is an initial period in which abnormalities in the region of the anterior temporal lobe undermine performance on semantic but not reading tests, followed by a second phase where these abnormalities extend to encompass areas responsible for direct lexical reading, followed by a third phase in which the neural areas that support nonlexical reading are also compromised. Coltheart et al. took great pains to point out that this spread-of-atrophy account, derived from the speculations of Noble, Glosser, and Grossman (2000), does not represent a prediction from the DRC model as "the model claims nothing about the neuroanatomical locations of its processing components, so cannot make such predictions" (Coltheart et al., 2010, p. 260). This disclaimer essentially renders the DRC model unfalsifiable on the basis of any data concerning structural or functional neural abnormalities in SD. Nevertheless, Coltheart et al. went on to suggest that a region likely to support direct lexical reading is an area of occipitotemporal cortex corresponding to the posterior portion of the left fusiform gyrus ($x = 43$, $y = -54$, $z = -12$, BA37) identified as the visual word form area by McCandliss, Cohen, and Dehaene (2003), and hence they "might expect abnormality of this region to result in surface dyslexia." (Coltheart et al., 2010, p. 260).

Yet, for at least some of our SD patients, surface dyslexic reading was apparent despite no evidence of structural abnormality in the region of BA37. As noted by Coltheart et al. (2010), two patients in our study, E.K. and B.S., were scanned longitudinally by Bright, Stamatakis, and Tyler (2008). E.K.'s reading performance in June of 2002 showed a surface dyslexic pattern (high-frequency regular words [HR] = 83%, low-frequency regular words [LR] = 48%, high-frequency exception words [HE] = 62%, low-frequency exception words [LE] = 33%) despite an absence of atrophy in BA37 on a scan in July of 2002 (Bright et al., 2008). Similarly, B.S.'s reading performance was assessed in November of 2002 and showed a rather pure surface dyslexic pattern (HR = 95%, LR = 91%, HE = 86%, LE = 50%), despite an absence of atrophy in BA37 both in the immediately preceding scan in July of 2002 and on the subsequent scan in October of 2003 (Bright et al., 2008). These data clearly demonstrate that structural damage to BA37 is not prerequisite for surface dyslexia in SD.

Nor does it seem that surface dyslexia in SD can be attributed to functional abnormality in the region of BA37. Coltheart et al. (2010) cited Nestor, Fryer, and Hodges (2006) as finding left fusiform hypometabolism in a group of SD patients. Coltheart et al. apparently regard the entire fusiform gyrus, which traverses the full length of the temporal lobe from back to front, as a single anatomical and functional region. It is not, and Nestor et al. explicitly stated that hypometabolism did not extend as far back as BA37 in the SD group. Moreover, eight of the nine SD cases reported in Nestor et al. were also in our cohort, and seven of these were scanned either in the same
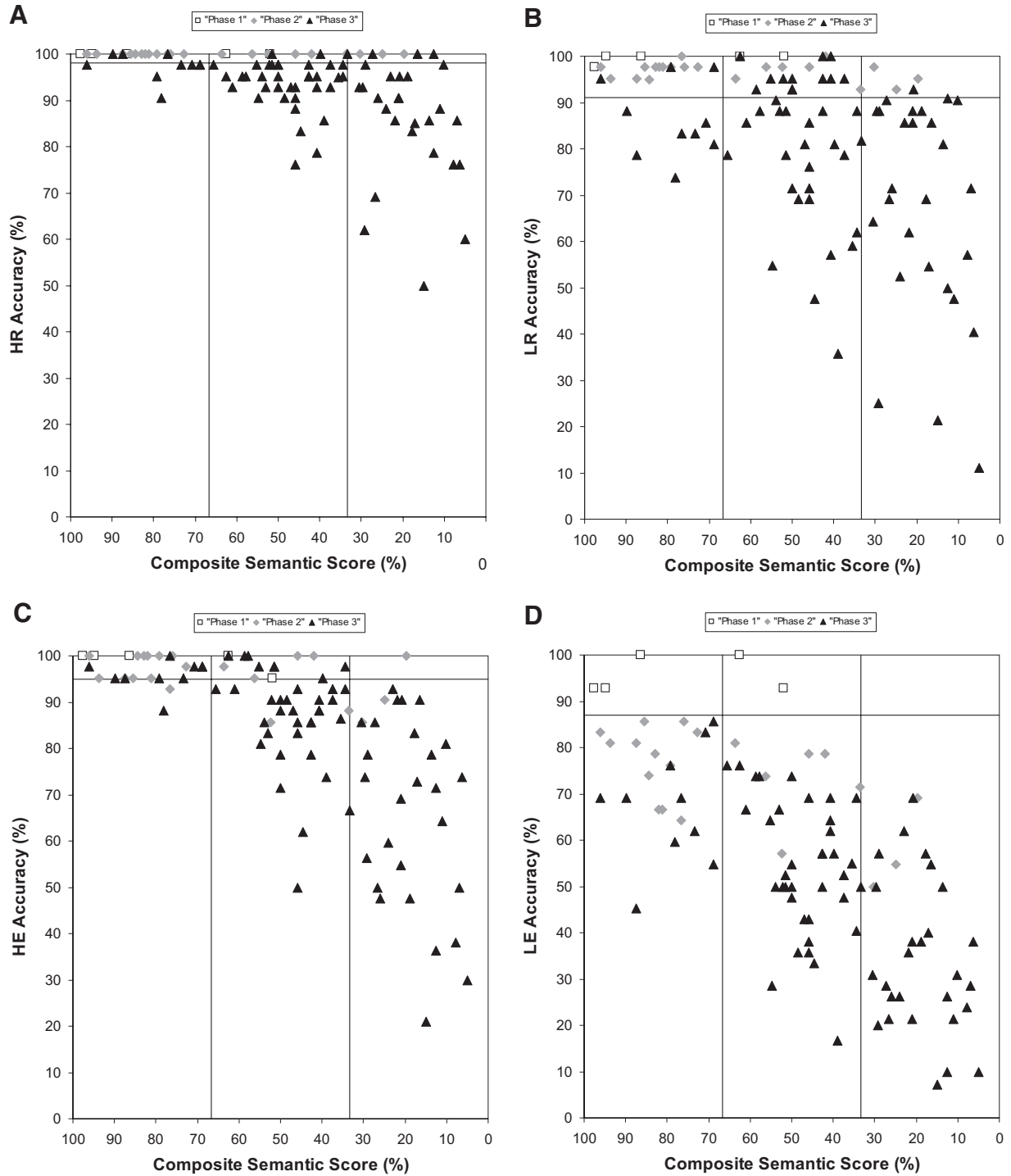
A

B

C

D

Figure 1. Semantic dementia patients' reading accuracy for each condition of the Surface List. A: High-frequency regular (HR) words. B: Low-frequency regular (LR) words. C: High-frequency exception (HE) words. D: Low-frequency exception (LE) words. The vertical lines represent scores of 66% and 33% on the composite semantic score, in an attempt to divide the observations into three levels of severity. The horizontal lines represent two standard deviations below control reading performance on any given condition. The symbols represent which of the purported phases each observation would fall into according to the Coltheart, Tree, and Saunders (2010) classification.

year as or substantially after collection of the reading data we reported. As this study did not find reliable hypometabolism in the posterior left fusiform region identified by Coltheart et al. as a likely seat for lexical processing and given that all seven of the cases were clearly surface dyslexic at or before the time of scanning, we can conclude that functional abnormalities in the region of BA37 did not cause their surface dyslexia.

Prevalence of a Nonword Reading Deficit in Late-Stage SD

In our patient data, we found nonword reading ability to be slightly impaired overall, averaging 78.53% correct ($SD = 22.51$) for the 34 occasions where it was assessed. Nonword reading ability was not, however, significantly predicted by the patients' composite semantic score, which accounted for only 3% of the available variance. Coltheart et al. (2010) simulated both the overall nonword reading impairment and the lack of relationship to the patients' semantic score. Yet, as mentioned earlier, Coltheart et al. subscribe to three phases of semantic dementia in which reading of all word classes is initially intact in Phase 1 but reading of all words classes is impaired by Phase 3, meaning that their account must predict some relationship between level of semantic knowledge and nonword reading ability. To demonstrate the validity of this three-phase prediction, Coltheart et al. then proceeded to remove five observations of nonword reading data from one patient (M.A.), after which composite semantic score became a (just) significant predictor of nonword reading performance in both patients and model, accounting for 12% and 15% of the available variance, respectively. Coltheart et al. offered this result as preliminary evidence that nonword reading deficits represent a general feature of late-stage SD.

There are a number of problems apparent in Coltheart et al.'s (2010) treatment of the relationship between nonword reading and level of semantic knowledge. First, the failure to find any evidence of such a relationship in the full set of 34 observations further undermines the plausibility of a distinct trichotomy of phases of reading in SD. Second, the justification for the removal of M.A.'s data offered by Coltheart et al. was that she constituted an outlier because her nonword reading deficit was apparent much earlier in the course of disease progression than all other cases tested on nonword reading. In this regard, it is worth noting that two of the three cases of preserved exception word reading in the face of an appreciable semantic deficit in our original data set were also formally outliers, but neither we nor Coltheart et al. considered this a sufficient justification for their removal from analysis despite the fact that this would have considerably increased the predictive capacity of the composite semantic score. Third, Coltheart et al. "have nothing to say about what might explain why M.A.'s nonword reading impairment became apparent at an earlier severity stage than is characteristic of the other patients" (Coltheart et al., 2010, p. 265). In contrast, we view outliers in our data to represent something to be explained rather than excluded. For all of these reasons, we consider the removal of M.A.'s nonword reading data to be unjustified and hence maintain that nonword reading ability was not significantly related to composite semantic score in our patient data. Consequently, there is currently no support for the proposal that a nonword reading deficit constitutes a general feature of late-stage SD.

Assessment of Model Fit: Comparing Apples and Oranges

Turning to the simulation data provided by Coltheart et al. (2010), their central claim is that the DRC model provides a better fit to the patient data than the connectionist triangle model simulation that we provided in terms of the numeric similarity of the intercepts and R^2 values for the regression of reading performance upon level of semantic knowledge. In fact, this assertion is unsustainable when one considers that the values provided for the DRC simulations are not comparable to those provided for the triangle model simulations. Coltheart et al. chose to use the SD patients' composite semantic scores to predict the reading accuracy of the DRC model for each condition. This choice is particularly surprising given that, as Coltheart et al. explicitly acknowledged, the DRC model has no semantic system to be lesioned because this source of knowledge is considered superfluous with respect to reading aloud (Coltheart et al., 2010, p. 259). Hence, the independent variable entering the regression analyses was derived from the SD patients, and the dependent variable was derived from the DRC model. In our simulations, the predictor of the reading accuracy of the model for each condition was the level of remaining "S"→P activation. In other words, both the independent and dependent variables entering the analyses were derived from the triangle model. Put simply, the measures of fit provided by Coltheart et al. are not comparable as the analyses for the DRC model used a patient variable to predict simulation data, whereas the analyses for the triangle model used a lesion parameter to predict simulation data.

A more valid comparison of the simulation performance of the two models would therefore be provided by considering the fit obtained when the DRC model's reading accuracy is regressed upon the parameters adjusted to provide these data, namely, the percentage of lexical entries and/or nonlexical rules removed to produce the pattern corresponding to any given observation. Table 1 provides the results of such an analysis, with intercepts and R^2 values given for regressions involving degree of lexical and nonlexical damage as predictors of DRC's reading performance, along with standardized slopes for each independent variable that indicate the extent to which each type of damage uniquely contributes to overall fit. As would be expected given the functional independence of the two pathways within the DRC model, both lexical damage and nonlexical damage are necessary to produce regular word reading deficits, whereas only lexical damage is required to produce exception word reading deficits and only nonlexical damage is needed to produce nonword reading deficits.

How does the fit of the DRC model to the patient data fare against that of the triangle model when evaluated on this level playing field? Table 2 provides the intercepts and R^2 values for the regression of reading accuracy for each condition on composite semantic score for the SD patients, along with the same measures for the original triangle model simulations we reported, plus additional values for the nonword condition that we consider in more detail shortly. We can then compare these to the values obtained when percentage lexical damage and percentage nonlexical damage are used as predictors of the reading behavior of the DRC model. This reanalysis shows that Coltheart et al.'s (2010) simulation data are in fact numerically less similar to the patient data than those of the triangle model for every condition on both measures. Hence, by Coltheart et al.'s own best fit criterion for theory evaluation, it would appear that the

Table 1

Intercepts and R^2 Values for Each Word Class for Regressions of the Dual Route Cascaded Model's Reading Accuracy Upon Level of Lexical and Nonlexical Damage, With Standardized Slopes for Each Variable

Word class	Intercept	R^2	Lexical slope	Nonlexical slope
High-frequency regular words	**112.9**	**0.47**	**−0.469**	**−0.301**
Low-frequency regular words	**123.8**	**0.59**	**−0.307**	**−0.556**
High-frequency exception words	**127.4**	**0.66**	**−0.848**	0.073
Low-frequency exception words	**122.3**	**0.90**	**−0.935**	−0.025
Nonwords	**122.4**	**0.66**	0.004	**−0.812**

Note. Significant relationships are provided in bold.

triangle model should be favored over the DRC model in accounting for reading aloud in SD.

Semantic Involvement in Nonword Reading: A Triangle Model Simulation

In evaluating the relative worth of each theory, Coltheart et al. (2010) placed a great deal of weight on DRC's ability to simulate the depressed level of nonword reading observed in some of our SD patients, as we did not attempt to model this effect. As we have already discussed above, nonword reading performance was not related to the level of semantic deficit, in contrast to the prediction of Coltheart et al.'s three-phase model. We have also demonstrated that nonword reading ability was significantly related nonlexical

Table 2

Intercepts and R^2 Values for Each Word Class for Regressions of Patients' Reading Accuracy on Composite Semantic Score, the Triangle Model's Reading Accuracy Upon Amount of Remaining Semantic Activation of Phonology, and the DRC Model's Reading Accuracy Upon Level of Lexical and Nonlexical Damage

Word class	Intercept	R^2
High-frequency regular words		
Patient data	**84.5**	**0.26**
Triangle model	**87.7**	**0.43**
DRC model	**112.9**	**0.47**
Low-frequency regular words		
Patient data	**63.7**	**0.26**
Triangle model	**60.5**	**0.54**
DRC model	**123.8**	**0.59**
High-frequency exception words		
Patient data	**62.9**	**0.42**
Triangle model	**59.1**	**0.57**
DRC model	**127.4**	**0.66**
Low-frequency exception words		
Patient data	**26.8**	**0.50**
Triangle model	**24.6**	**0.84**
DRC model	**122.3**	**0.90**
Nonwords		
Patient data	71.4	0.03
Triangle model	*80.8*	*0.00*
DRC model	**122.4**	**0.66**

Note. Significant relationships are provided in bold. The italicized entry for nonwords for the triangle model refers to values from the new simulation provided in this article. DRC = dual route cascaded.

lesion severity in the DRC simulations. We would suggest that, in light of these facts, the DRC model cannot be considered to have successfully simulated nonword reading performance in SD. Nevertheless, it is true that we did not attempt to simulate the patients' nonword reading data in our original article. Following Plaut et al. (1996), we assumed that nonwords do not elicit a significant degree of semantic activation and hence would not be affected by the corruption of "S"→P activation due to lesioning.

There is, however, reason to expect that semantic damage might have a mild, generalized impact on nonword reading. Within models that incorporate O→S connections (e.g., Harm & Seidenberg, 2004), we would expect partial activation of the semantic representations of a nonword's orthographic neighbors, which would presumably then flow on to phonology. We proposed that in SD, semantic degradation means that the semantic activation propagating to phonology is both reduced and distorted. Reduction of this activation would be expected to have a minimal impact on nonword reading, given that the correct pronunciation for such items must ultimately be driven by activation from the direct pathway; but the perturbation of the low-level "S"→P signal elicited by these items should nonetheless have a generally negative impact. Indeed, when we implemented this hypothesis within the model used in Woollams et al. (2007) through addition of a small constant amount of Gaussian noise ($SD = 0.85$) to phonology for nonwords, the simulation produced an average nonword reading level comparable to that of the patients (81.4%, range: 49%–100%, and 78.9%, range: 23%–100%, respectively). The variation in nonword reading accuracy at any given level of semantic deficit corresponds to degree of premorbid division of labor for that network, such that the greater the semantic support provided during training, the more vulnerable nonword reading performance is to noise at lesioning. This simulation therefore approximates the considerable individual differences in nonword reading accuracy observed amongst the SD patients we considered. As can be seen in both Table 2 and in comparison of Figures 2A and 2B, the simulation data corresponds to the data from SD patients in that neither reveals a significant relationship between nonword reading accuracy and level of the semantic deficit.

Predicting Regular Word Reading: A Theory-Independent Capacity

Another reason that Coltheart et al. (2010) asserted the superiority of the DRC account of reading aloud in SD over the triangle model account relates to the predictive capacity of what they termed the "dual route equation" (p. 265). This equation—
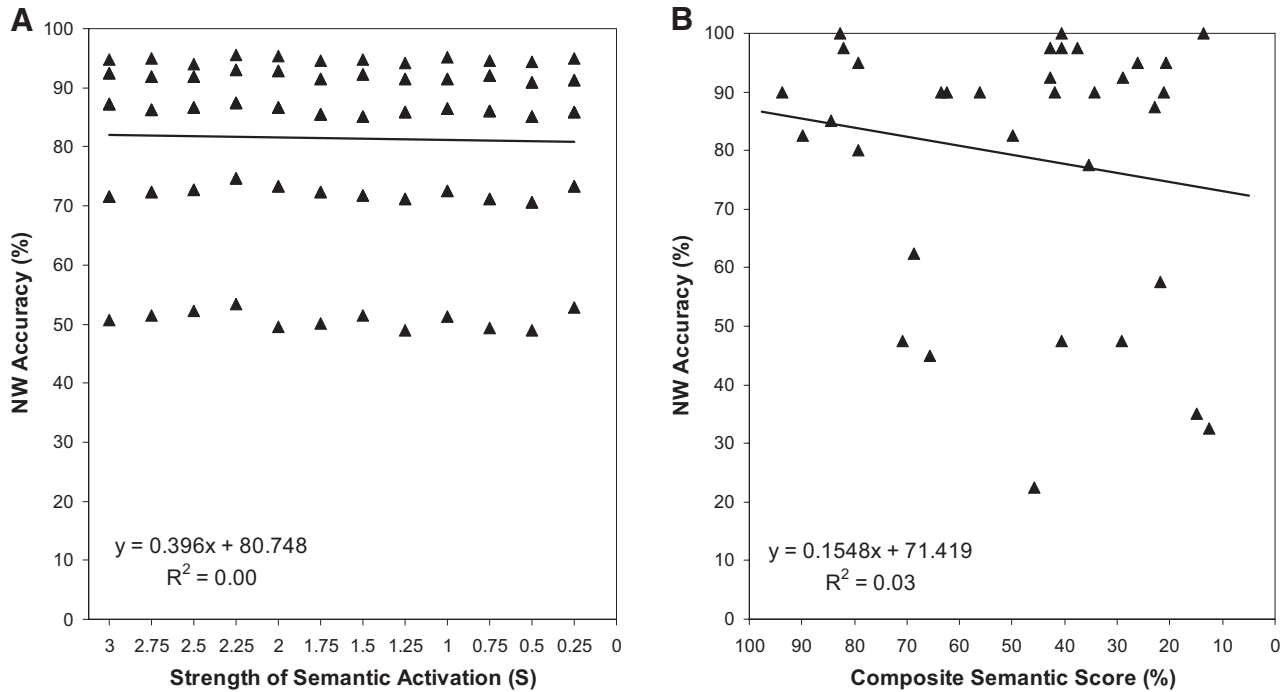
Figure 2. Reading accuracy for the Surface List nonwords according to severity of semantic deficit for (A) the connectionist triangle model with a constant level of noise applied at lesioning and (B) the 34 occasions in which this ability was assessed in the semantic dementia patients. NW = nonword.

pr(regular word correct) = pr(exception word correct) + (1 − pr[exception word correct]) × pr(nonword correct)—allows generation of a predicted level of regular word reading accuracy on the basis of an individual's exception word and nonword reading performance. As noted by Coltheart et al., these predicted values correspond closely to those observed amongst the SD patients we reported, with strong correlations for both high- and low-frequency items (.928 and .924, respectively). It is worth stressing at this point that there is no sense in which this equation is specific to the DRC model: It embodies a general prediction concordant with any model that incorporates at least partially independent procedures for reading exception words and nonwords. As such, the predictive capacity of the DRC equation in the patient data is simply a fact about the patients' reading performance that should be reflected in the simulation data of any adequate model. When we applied this equation to the 34 observations of word reading data that also included nonword reading scores from the DRC simulations, we obtained strong correlations between predicted and observed regular word reading accuracy for both high- and low-frequency items (.949 and .918, respectively). Critically, when we applied this equation to the triangle model data, which constituted 108 word reading scores from our original simulation and the 108 nonword reading scores from the simulation provided earlier and presented in Figure 2, we obtained similarly strong correlations for both high- and low-frequency items (.938 and .905, respectively). There seems to be little that is uniquely dual route about this particular equation, and hence, its accurate prediction of the SD patients' regular word reading does not bear on the relative theoretical adequacy of the DRC versus triangle model accounts of reading aloud in SD.

Theory Evaluation

The Goal of Computational Modeling: Explanation, Not Emulation

As discussed at length by Seidenberg and Plaut (2006), the DRC and connectionist triangle accounts represent two contrasting approaches to the computational modeling of reading, with differing goals as well as methods. As a result of the discrepancy in goals, the advocates of each account use rather different sets of criteria to evaluate theoretical adequacy. The primary goal of the DRC approach is to fit as much data as closely as possible, and hence, its proponents consider the model that achieves this goal to be the superior one. In their article, Coltheart et al. (2010) followed this logic in asserting that the DRC model provides a better account of reading aloud in SD than our connectionist triangle model. As we have already argued, this conclusion is not valid when one considers analyses of fit that are comparable across the two models. In fact, the triangle model simulations are to be preferred on the criteria of fit to the SD patient reading data alone. Nonetheless, it is not on the basis of a closer fit to the target data that we consider our account of reading aloud in SD to be preferable to that of DRC.

The goal of computational modeling is surely to explain the patterns of performance seen in human behavior in some cognitive domain. We do not deny that Coltheart et al. (2010) were reasonably successful at emulating the patient data through selecting the closest match for each observation from a very large pool. Yet this approach offers no insight into the underlying cognitive or neural bases of the reading disorder. As we argued in our original article, it does not explain why there is such a strong association between

level of semantic knowledge and exception word reading accuracy. In dismissing this relationship as an accident of neuroanatomical contiguity (a view not supported by the evidence we considered earlier), Coltheart et al. thereby failed to explain the vast majority of the SD reading aloud data that we presented.

Fitting the Noise: Implausible Longitudinal Profiles of Lesion Severity in SD

In addition to lacking explanatory capacity, the bottom-up data-driven approach to computational modeling associated with DRC can have undesirable consequences. In essence, when fitting the data represents the gold standard for evaluating a theory's worth, the drive to emulate the target data precisely means that not only is the signal of interest simulated but also the noise in measurement. The Coltheart et al. (2010) simulations provide a prime example of the perils associated with such overfitting. There is no denying that SD represents a progressive disease, in terms of both semantic degradation and the atrophy that underpins it (Rohrer et al., 2008). Yet, when we consider the longitudinal profiles of lexical and nonlexical damage in the simulations of Coltheart et al., we see that increasing disease severity was accompanied by a decrease in the amount of lexical damage applied in 10.20% of cases and a decrease in the amount of nonlexical damage applied in 26.53% of cases. To take one example, in fitting the data from Patient V.H. across Testing Rounds 3, 5, and 7, the degree of lexical damage was assumed to decrease from 45% to 34.5% before increasing again to 46%, while the degree of nonlexical damage steadily decreased from 47.5% to 20% to 9.5%.

If such implausible decreases in level of damage are required to obtain effectively identical fits to those that we produced with premorbid variation and continuous degradation of a single theoretically motivated parameter (semantic support of phonology), then surely these DRC simulations cannot be regarded as providing a superior account of the patient data. Coltheart et al. (2010) regarded success in precise simulations of individual patient data as a major virtue of their approach, yet we have shown that this strategy comes at the cost of overfitting, which undermines the plausibility of the DRC account. Of course, we could have varied every possible combination of training and lesioning parameters within our model to produce a huge number of data sets from which we then selected the best fitting observation for each individual patient. Our reason for eschewing this strategy is that such overfitting actually works against a fuller theoretical understanding of the critical variables that underpin the reading deficits observed in SD, as apparent in the implausible longitudinal profiles of lexical and nonlexical damage produced by the DRC model.

Defining Parsimony: Parameters Varied or Deficits Explained?

In our original article, we suggested that our account of surface dyslexia in SD represented a parsimonious explanation. One way that parsimony may be defined involves the number of parameters that are required to simulate a particular set of target phenomena within any given domain. Coltheart et al. (2010) suggested that in our simulations, variation in two parameters was required: the extent of "S"→P activation provided during training and the

reduction of "S"→P activation during lesioning. They considered this to be comparable to the two parameters varied in their simulations, namely, the extent of lexical and nonlexical damage. We note, however, that the central feature of SD, namely, the multimodal semantic deficit, is not captured by the Coltheart et al. simulation; should any future version of the DRC model incorporate a working semantic system, an extra parameter would be required to capture this aspect of the patients' performance.

We would not, however, claim that our triangle model account of the SD-squared phenomenon offers a more parsimonious explanation than DRC purely on the basis of the number of parameters varied in each simulation. We subscribe to the view expressed by Seidenberg and Plaut (2006) that the theoretical adequacy of these contrasting approaches must also be assessed according to the extent to which they engage with other aspects of cognition. The simulations we provided represent an instantiation of a more general notion common to the connectionist approach, which is that amodal semantic knowledge supports processing of atypical items. Patterson et al. (2006) demonstrated that, far from just capturing the performance of SD patients in reading aloud, this hypothesis predicts the conspicuous Frequency × Typicality interaction seen across a diverse set of productive and receptive tasks involving verbal (spelling, past-tense generation, lexical decision) and nonverbal (delayed copy drawing and object decision) materials. Hence, we consider the triangle model account of reading aloud in SD to be more parsimonious than that offered by DRC on the basis of its applicability across multiple different cognitive domains.

Convergent Measures and Novel Predictions

Another criterion identified by Seidenberg and Plaut (2006) for theory evaluation is the extent to which an account generates novel empirical predictions. The DRC simulation of the SD-squared data involved manipulation of two parameters in the form of degree of lexical and nonlexical damage. Yet there does not seem to be any independent nonreading behavioral measure that could be used to tap the integrity of either of these forms of processing. Nor does the DRC account make any predictions as to the neuroanatomical bases of these abilities. In the absence of any independent measures of the key parameters underpinning the DRC account, it is not immediately apparent what novel empirical predictions are offered by the simulations reported by Coltheart et al. (2010).

In contrast, the connectionist triangle account proposes that reading performance is linked to other cognitive functions (Patterson & Lambon Ralph, 1999). This primary systems view means that the connectionist account of reading disorders makes clear predictions concerning performance on independent nonreading behavioral measures. In the case of surface dyslexia, exception word reading deficits should be paralleled by similarly impaired performance on any task that assesses the integrity of amodal semantic representations, and this is undeniably true of the patient data we reported. We modeled outlying cases in this distribution of reading impairment through assuming variations in the degree of premorbid division of labor. Although difficult to validate retrospectively in patient populations, this account offers the clear prediction that there should be a relationship between the magnitude of regularity and semantic effects amongst normal readers, such that the degree of interaction between spelling–sound typicality and imageability (Shibahara, Zorzi, Hill, Wydell, & Butterworth, 2003; Strain & Herdman, 1999; Strain, Patterson, & Seidenberg,

1995, 2002; Woollams, 2005) may be used as an index of division of labor.

Conclusions

In this reply, we have argued that our connectionist triangle model account of reading aloud in SD remains superior to the DRC model account offered by Coltheart et al. (2010). Our account is entirely consistent with what is currently known about the neuroanatomy of reading aloud in SD, and our simulations provide, if anything, a better fit to the SD patients' reading aloud data that we reported. More importantly, our account parsimoniously explains not only these patients' semantic and reading impairments but also their deficits in processing atypical items across multiple domains. In addition, our hypotheses concerning individual differences in premorbid division of labor have generated novel empirical predictions concerning normal reading that we are in the process of investigating. Hence, we conclude that on all fronts, the connectionist triangle model account continues to provide the most faithful and fruitful explanation of the SD-squared phenomenon.

References

Brambati, S. M., Ogar, J., Nehaus, J., Miller, B. L., & Gorno-Tempini, M. L. (2009). Reading disorders in primary progressive aphasia: A behavioral and neuroimaging study. *Neuropsychologia, 47,* 1893–1900.

Bright, P., Moss, H. E., Stamatakis, E. A., & Tyler, L. K. (2008). Longitudinal studies of semantic dementia: The relationship between structural and functional changes over time. *Neuropsychologia, 46,* 2177–2188.

Coltheart, M. (2006). Acquired dyslexias and the computational modelling of reading. *Cognitive Neuropsychology, 23,* 96–109.

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.

Coltheart, M., Tree, J. J., & Saunders, S. J. (2010). Computational modeling of reading in semantic dementia: Comment on Woollams, Lambon Ralph, Plaut, and Patterson (2007). *Psychological Review, 117,* 256–272.

Fushimi, T., Komori, K., Ikeda, M., Lambon Ralph, M. A., & Patterson, K. (2009). The association between semantic dementia and surface dyslexia in Japanese. *Neuropsychologia, 47,* 1061–1068.

Graham, N., Patterson, K., & Hodges, J. R. (2000). The impact of semantic memory impairment on spelling: Evidence from semantic dementia. *Neuropsychologia, 38,* 143–163.

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.

Hodges, J. R., Graham, N., & Patterson, K. (1995). Charting the progression in semantic dementia: Implications for the organisation of semantic memory. *Memory, 3,* 463–495.

Jefferies, E., Lambon Ralph, M. A., Jones, R., Bateman, D., & Patterson, K. (2004). Surface dyslexia in semantic dementia: A comparison of the influence of consistency and regularity. *Neurocase, 10,* 290–299.

Lambon Ralph, M. A., Graham, K. S., Ellis, A. W., & Hodges, J. R. (1998). Naming in semantic dementia: What matters? *Neuropsychologia, 36,* 775–784.

McCandliss, B. D., Cohen, L., & Dehaene, S. (2003). The visual word form area: Expertise for reading in the fusiform gyrus. *Trends in Cognitive Science, 7,* 293–299.

Mummery, C. J., Patterson, K., Price, C. J., Ashburner, J., Frackowiak, R. S. J., & Hodges, J. R. (2000). A voxel-based morphometry study of semantic dementia: Relationship between temporal lobe atrophy and semantic memory. *Annals of Neurology, 47,* 36–45.

Nestor, P. J., Fryer, T. D., & Hodges, J. R. (2006). Declarative memory impairments in Alzheimer's disease and semantic dementia. *NeuroImage, 30,* 1010–1020.

Noble, K., Glosser, G., & Grossman, M. (2000). Oral reading in dementia. *Brain and Language, 74,* 48–69.

Patterson, K., & Hodges, J. R. (1992). Deterioration of word meaning: Implications for reading. *Neuropsychologia, 30,* 1025–1040.

Patterson, K., & Lambon Ralph, M. A. (1999). Selective disorders of reading? *Current Opinion in Neurobiology, 9,* 235–239.

Patterson, K., Lambon Ralph, M. A., Jefferies, E., Woollams, A., Jones, R., Hodges, J. R., & Rogers, T. T. (2006). "Presemantic" cognition in semantic dementia: Six deficits in search of an explanation. *Journal of Cognitive Neuroscience, 18,* 169–183.

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.

Rogers, T. T., Lambon Ralph, M. A., Garrard, P., Bozeat, S., McClelland, J. L., Hodges, J. R., & Patterson, K. (2004). Structure and deterioration of semantic memory: A neuropsychological and computational investigation. *Psychological Review, 111,* 205–235.

Rohrer, J. D., McNaught, E., Foster, J., Clegg, S. L., Barnes, J., Omar, R., . . . Fox, N. C. (2008). Tracking progression in frontotemporal lobar degeneration: Serial MRI in semantic dementia, *Neurology, 71,* 1445–1451.

Seidenberg, M. S., & Plaut, D. C. (2006). Progress in understanding word reading: Data fitting versus theory building. In S. Andrews (Ed.), *From inkmarks to ideas: Current issues in lexical processing* (pp. 25–49). Hove, England: Psychology Press.

Shibahara, N., Zorzi, M., Hill, M. P., Wydell, T., & Butterworth, B. (2003). Semantic effects in word naming: Evidence from English and Japanese kanji. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 56*(A), 263–286.

Strain, E., & Herdman, C. M. (1999). Imageability effects in word naming: An individual differences analysis. *Canadian Journal of Experimental Psychology, 53,* 347–359.

Strain, E., Patterson, K., & Seidenberg, M. S. (1995). Semantic effects in single-word naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 1140–1154.

Strain, E., Patterson, K., & Seidenberg, M. S. (2002). Theories of word naming interact with spelling–sound consistency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 207–214.

Woollams, A. M. (2005). Imageability and ambiguity effects in speeded naming: Convergence and divergence. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(5), 878–890.

Woollams, A. M., Cooper-Pye, E., Hodges, J. R., & Patterson, K. (2008). Anomia: A doubly typical signature of semantic dementia. *Neuropsychologia, 46,* 2503–2514.

Woollams, A. M., Lambon Ralph, M. A., Plaut, D. C., & Patterson, K. (2007). SD-squared: On the association between semantic dementia and surface dyslexia. *Psychological Review, 114,* 316–339.

Postscript: SD-Squared Revisited Again

Anna M. Woollams and Matthew A. Lambon Ralph
University of Manchester

David C. Plaut
Carnegie Mellon University

Karalyn Patterson
Medical Research Council Cognition and Brain Sciences Unit

We begin by responding to the final challenge posed by Coltheart, Tree, and Saunders (2010) in their postscript, which is, if there is no single, unified connectionist triangle model of reading forthcoming, "what is the point of the experimental psychology of reading, and what is it that computational modelers of reading are trying to model?" (p. 272). Our answer is that both experimental and computational methodologies should be aimed at explaining human behavior by uncovering the computational principles that govern the underlying cognitive and neural processes. While it is important to ensure that the explanation and principles remain self-consistent, there is no need for a single uber-simulation, just as there is no need for a single behavioral experiment that manipulates all relevant variables simultaneously. Detailed comparison between model and human performance is critical for evaluating and improving current understanding, but data fitting per se is not the goal of the enterprise.

Moreover, our approach differs from that of Coltheart et al. (2010) in terms of the type and range of data we consider most informative. When dealing with neuropsychological data, the connectionist triangle model approach attempts to capture the central tendency and distribution of a set of observations. Additionally, it pays individual attention to outliers from this distribution not only for the sake of completeness but also because, with the traditional emphasis of cognitive neuropsychology on single-case studies, such atypical patterns have often constituted the main text rather than a footnote (Patterson & Plaut, 2009). We should stress that our work on reading aloud in semantic dementia (SD) did not present averaged data of the kind usually seen in a group study (Woollams, Lambon Ralph, Plaut, & Patterson, 2007). Rather, we employed the case-series approach by presenting data for individual observations, which incorporates the virtues of the single-case study methodology along with an appropriate stress on the most typical pattern of data.

Coltheart et al. (2010) argued that the optimal level of analysis is instead that of the individual observation, which guides their approach to simulation of reading aloud in SD. Their simulations involved a huge pool of 40,200 observations generated by varying two independent lesion parameters; from these, they selected the subset most closely fitting the patient data we reported (Woollams et al., 2007). From our perspective, this approach means that noise in the data is modeled in addition to the signal of interest. This results in the occasional implausible longitudinal profile of damage for a given patient within the dual route cascaded (DRC) model simulations, such that, for example, increasing disease severity is accompanied by decreasing levels of lexical and/or nonlexical damage.

To assess the adequacy of the triangle model account, we compared the relationship between semantics and reading observed in the patient data to the relationship between "semantics" and reading seen in the simulation data. The parallel assessment for the DRC account is to compare the relationship between severity (as represented by composite semantic score) and reading observed in the patient data to the relationship between "severity" (as represented by degree of lexical and nonlexical damage) and reading seen in the simulation data. When we calculated these measures, both models succeeded reasonably well in capturing the patient results for all word classes. There was, however, a strong relationship between severity and nonword reading in the DRC model that was not apparent in either the patient or connectionist triangle model data. Although Coltheart et al. (2010) suggested that our simulation of nonword reading did not provide any explanation of the range of performance observed, it was clear from the data we presented that the degree of impairment was linked to the premorbid division of labor in the network, such that the more extreme the division of labor, the greater the nonword reading deficit observed after lesioning.

Another point worth clarifying concerns the nature of the spread of atrophy over the course of disease progression in SD. There is no doubt, as argued by Coltheart et al. (2010), that in SD—as with all progressive brain diseases—the pathology advances from its original focus, which in SD is the anterior, inferior temporal lobe. This means that, for patients with severe SD, areas adjacent to the temporal pole (posterior, anterior, or superior; not inferior because the earliest atrophy is already in the most inferior part of the brain) will also be abnormal. Because Coltheart et al. assumed that the surface dyslexia associated with SD is unrelated both to the patients' semantic disorder and to the anterior temporal lobe, they argued that

> the ability of patients to name pictures and match spoken words [i.e., their semantic status] is a measure of the extent of the patients' cortical atrophy and that the greater this atrophy is, the more damage there will be to nonsemantic components of the reading system. (Coltheart et al., 2010, p. 271)

We doubt that this claim can explain the fact that, as Coltheart et al. (2010) acknowledged, surface dyslexic reading is ultimately a universal feature of SD. There are several reasons for our doubt. First of all, the spread of atrophy in SD beyond the anterior temporal lobe is variable across patients in terms of both location and rate (Mummery et al., 2000; Rohrer et al., 2008). Why should a patient whose additional damage is rostral to the temporal pole (in the frontal lobe) and one for whom it is caudal (in the posterior temporal lobe) both be surface dyslexic? Even more seriously, as we emphasized in our original article, the surface dyslexia characteristic of SD patients is the manifestation in the domain of reading of a much more general phenomenon. These patients are strikingly and consistently impaired in processing less familiar and less typical items across domains as divergent as reading, spelling, verb inflection, lexical decision, object decision, delayed copy drawing, knowledge of object color, and more (e.g., Patterson et al., 2006; Rogers, Patterson, & Graham, 2007). In all of these tasks, as well as in tests tapping semantic memory more directly (such as object naming; e.g., Woollams, Cooper-Pye, Hodges, &

Patterson, 2008), the pattern of SD performance is the same: not only significantly reduced accuracy for less familiar, atypical items but errors in which the response is more typical of its domain than the stimulus. Because the connectionist triangle model proposes that semantic information underpins effective processing of infrequent atypical items in any domain, it predicts precisely this cross-domain pattern of impairment in SD. What is the DRC model's account of this pattern?

References

Coltheart, M., Tree, J. J., & Saunders, S. J. (2010). Computational modeling of reading in semantic dementia: Comment on Woollams, Lambon Ralph, Plaut, and Patterson (2007). *Psychological Review, 117,* 256–272.

Mummery, C. J., Patterson, K., Price, C. J., Ashburner, J., Frackowiak, R. S. J., & Hodges, J. R. (2000). A voxel-based morphometry study of semantic dementia: Relationship between temporal lobe atrophy and semantic memory. *Annals of Neurology, 47,* 36–45.

Patterson, K., Lambon Ralph, M. A., Jefferies, E., Woollams, A., Jones, R., Hodges, J. R., & Rogers, T. T. (2006). "Presemantic" cognition in semantic dementia: Six deficits in search of an explanation. *Journal of Cognitive Neuroscience, 18,* 169–183.

Patterson, K., & Plaut, D. C. (2009). "Shallow draughts intoxicate the brain": Lessons from cognitive science for cognitive neuropsychology. *Topics in Cognitive Science, 1,* 39–58.

Rogers, T. T., Patterson, K., & Graham, K. S. (2007). Colour knowledge in semantic dementia: It is not all black and white. *Neuropsychologia, 45,* 3285–3298.

Rohrer, J. D., McNaught, E., Foster, J., Clegg, S. L., Barnes, J., Omar, R., . . . Fox, N. C. (2008). Tracking progression in frontotemporal lobar degeneration: Serial MRI in semantic dementia, *Neurology, 71,* 1445–1451.

Woollams, A. M., Cooper-Pye, E., Hodges, J. R., & Patterson, K. (2008). Anomia: A doubly typical signature of semantic dementia. *Neuropsychologia, 46,* 2503–2514.

Woollams, A. M., Lambon Ralph, M. A., Plaut, D. C., & Patterson, K. (2007). SD-squared: On the association between semantic dementia and surface dyslexia. *Psychological Review, 114,* 316–339.