Available online at www.sciencedirect.com
SCIENCE DIRECT®
Cognitive Systems Research 6 (2005) 61–69
Cognitive Systems RESEARCH
www.elsevier.com/locate/cogsys

A computational exploration of double dissociations: modes of processing instead of components of processing

Action editor: Christian Schunn

Daragh E. Sibley *, Christopher T. Kello

Department of Psychology, George Mason University, Fairfax, VA 22030, USA

Received 16 September 2004; accepted 16 September 2004

Abstract

Two connectionist models are reported that simulated the defining features of the double dissociation between phonological and surface dyslexia in word reading. One model was a feed-forward, three-layer perceptron, and the other included recurrent connections. Neither model contained an architectural separation of sublexical and lexical processes, nor of phonological and semantic processes. Analyses showed that the double dissociation was simulated because the control parameter *input gain* shifted the models between conjunctive and componential modes of processing. The dissociation was not simulated by any kind of damage to separate system components. The simulations are discussed in the context of current accounts of surface and phonological dyslexia.

Keywords: Double dissociations; Connectionism; Control parameters; Dyslexia; Modularity; Input gain

1. Introduction

In most current models of word reading, two routes of processing are proposed to compute the sound of a word given its printed form (Coltheart, Curtis, Atkins, & Haller, 1993; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Harm & Seidenberg, 1999; Plaut, McClelland, Seidenberg, & Patterson, 1996). One route is *sublexical* in that it extracts and utilizes regularities in the mapping from components of orthography (e.g., letters) to components of phonology (e.g., phonemes). For instance, the letter P typically corresponds to the sound /p/, as in PIT. The other route is *lexical* in that it uses either localist or semantic representations of words. Localist and semantic representations are both considered lexical in this context because they bear little or no systematic relationship to the fine-grained components of orthography and phonology. For instance, the letter P and sound /p/ bear no systematic relationship to the meanings of the word PIT.

* Corresponding author.
E-mail address: dsibley@gmu.edu (D.E. Sibley).

1389-0417/$ - see front matter
doi:10.1016/j.cogsys.2004.09.007

The evidence for sublexical and lexical routes of processing in word naming has come from a number of sources (for a review, see Plaut et al. (1996)), but here, we focus on evidence from selective reading impairments that occur as a result of brain damage. One class of impairment termed *phonological dyslexia* is characterized by poor reading aloud of nonwords (e.g., SHONG), with relatively intact word reading (Behrmann & Bub, 1992). Conversely, *surface dyslexia* is characterized by poor reading aloud of words with irregular spelling-sound correspondences, with relatively intact nonword reading (Funnell, 1983). For instance, the vowel correspondence in PINT is irregular because the letter I, as well as the body INT, typically correspond to a short vowel, rather than the long vowel in PINT.

The complementary impairments of phonological and surface dyslexia have a straightforward explanation in terms of separate sublexical and lexical routes of processing. The impairment in nonword reading that characterizes phonological dyslexia arises from damage to the sublexical route. This is because performance on nonword stimuli presumably relies on the componential regularities between spelling and sound (e.g., the sounds corresponding to SH, O, and NG can be pieced together to form a plausible nonword pronunciation). By contrast, the impairment in irregular word reading that characterizes surface dyslexia arises from damage to the lexical route. This is because performance on irregular word stimuli presumably relies on word-specific knowledge (e.g., one must know the word PINT in order to pronounce it correctly).

There are a number of reasons why sublexical and lexical routes have been proposed to account for surface and phonological dyslexia, but perhaps the most important one is the basic logic of *double dissociations*. If two components of a cognitive system operate independently, then they will make independent contributions to the behavior of that system. Thus, a change in one component will have no effect on the behavioral contributions of the other. Such complementary effects on behavior constitute a double dissociation, and they are often interpreted as strong evidence that independent system components underlie the dissociated behaviors.

Surface and phonological dyslexia constitute a double dissociation, and many researchers have interpreted this double dissociation as evidence for sublexical and lexical routes. Some researchers have alternatively proposed that semantic and phonological components may underlie this dissociation (Patterson & Ralph, 1999), but this alternative also assumes that two system components underlie the dissociation, as in the dual-route hypothesis.

Despite appearances, even the purest cases of phonological and surface dyslexia do not necessitate the existence of two processing routes or two system components of any kind (Plaut, 1995; Van Orden, Pennington, & Stone, 2001). A single-component explanation is always a logical possibility, but without a specific single-component account, it is just a logical possibility. Until such accounts are proposed, observed dissociations will continue to be interpreted as evidence for the existence of two system components in word reading.

Recently, Kello and Plaut (2003) reported on a model of word reading that offers a true single-component explanation of the double dissociation between phonological and surface dyslexia. The model was inspired by the basic question of how reading acquisition builds upon the prior learning that occurs during spoken language acquisition (also see Plaut & Kello (1999)). The acquisition of spoken language requires the mapping from sound to meaning (comprehension) and meaning to sound (production). In the context of a connectionist approach to lexical processing, a single, distributed level of representation can be learned to mediate the bi-directional mapping between the phonological and semantic attributes of words. The question on this approach is, how do printed word forms make contact with the bi-directional mapping learned during spoken language acquisition?

The answer offered by Kello and Plaut (2003) is that orthography maps into the level of representation that mediates semantics and phonology (see Fig. 1, inset), rather than mapping into semantics and phonology themselves. This architecture effectively "kills two birds with one stone": the mediating representations provide access to both the semantic and phonological forms, so only one route is necessary. While there is a logical appeal

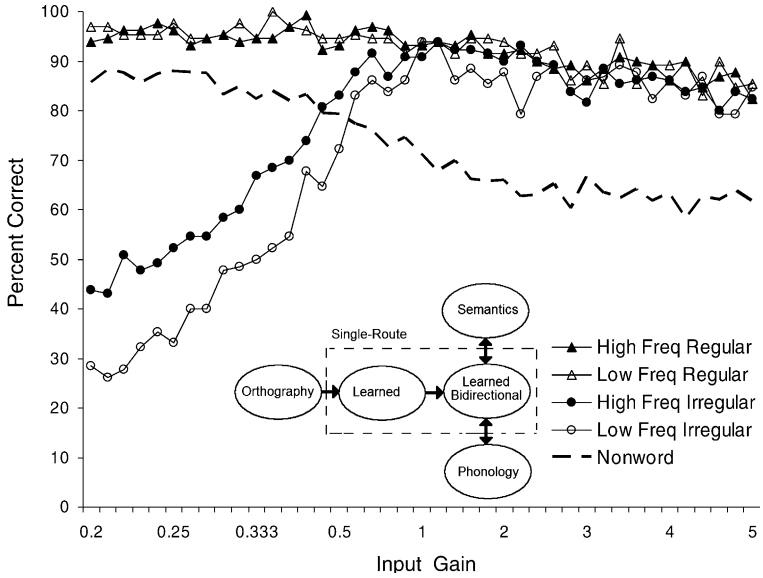

Fig. 1. Results from the single-route model of word reading reported by Kello (2003).

to this single-route architecture, it is unclear how it could account for the dissociation between surface and phonological dyslexia because there is no architectural separation of sublexical and lexical processes.

In a subsequent analysis of this single-route model, Kello (2003) showed that an impairment in the processing of nonwords could, in fact, be dissociated from an impairment in the processing of irregular words. The method used to simulate this dissociation was to shift the model between sublexical and lexical "modes" of processing by means of a control parameter termed *input gain* (see Section 2.2). The simulation results showing the dissociation are graphed in Fig. 1. The use of input gain was inspired by simulations of other behavioral data (see Kello & Plaut (2003)), but the function of input gain such that it caused this dissociation was not elucidated.

2. Current simulations

The goal of the current work was to use computational analyses to determine how input gain had its dissociating effect in the simulation of word reading reported by Kello (2003). To facilitate these analyses, the essential principles and mechanisms of the single-route model of word reading were distilled into two very simple connectionist models: one model was a feed-forward, 3-layer perceptron, and the other included recurrent connections from the output layer back to the hidden layer. These models did not simulate word reading or dyslexia, although their results are compared with data from experiments with phonological and surface dyslexics. Rather, the main purpose of the models was to replicate the dissociating effect of input gain in a more controlled and analytically tractable context. This tractability enabled us to determine that input gain can shift a connectionist model between *componential* and *conjunctive* modes of processing, and that it is this function of input gain that was responsible for the simulation of phonological and surface dyslexia reported by Kello (2003).

2.1. Input and output representations

Input and output representations were constructed from a 12 dimensional binary space. Out of $2^{12} = 4096$ possible input patterns, one fourth (1024) were chosen at random to constitute the

corpus of known items. Each chosen input pattern was associated with one output pattern. Output patterns were created in three steps. First, each input pattern was copied to its corresponding output pattern (i.e., the identity mapping). Second, frequencies were assigned to each item according to Zipf's law, $f = r^{-0.5}$, where r was an arbitrarily assigned rank from 1 to 1024. Third, the bit value of each dimension, for each output pattern, was flipped with a probability governed by Zipf's law, $p = 0.82r^{-0.5}$. The result of this formula was that the more frequent items were more likely to be irregular, and more likely to be more irregular (i.e., have more flipped values), compared with the less frequent items. This formula characterizes the relationship between frequency and regularity that exists in the English language, as well as other languages. The multiplicative constant of 0.82 was set such that there was a 5% probability on average of flipping each target value across the set of known items. There were 580 fully regular items (no flipped bits), and 444 irregular items with at least one flipped bits per item. The 3072 remaining patterns served to test the generalization of learning to novel items.

Each of the 12 input dimensions were coded by two input units, one coding on-bits as 1 and off-bits as 0, the other coding the opposite. This $x|1-x$ coding scheme was used because the models were trained via backpropagation. In backpropagation, no learning will occur on a unit's sending weights when the activation value of that unit is zero. Therefore, the $x|1-x$ coding scheme ensured that weight derivatives were generated for every input dimension, on every training episode. The $x|1-x$ coding was not necessary for the output units, so there were only 12 output units, each one corresponding directly to one of the 12 output dimensions.

The input and output representations captured the essential properties of quasi-regularity as it is implemented in most connectionist models of word reading. Specifically, each input unit had a systematic relationship with one output unit, much like the way that each orthographic unit would have a systematic relationship with at least one phonological unit (e.g., a unit for the letter P in the initial position would have a systematic relationship with a unit for the phoneme /p/ in the first position). Moreover, these relationships were never entirely systematic, much like real quasi-regularity in spelling-sound correspondences.

2.2. Model architecture

In both the feed-forward model and the recurrent model, the input units were fully connected to 200 hidden units, and the hidden units were fully connected to the output units. In the recurrent model, the output layer was also fully connected back to the hidden layer. The number of hidden units was determined through pilot testing to be about 50 units more than the minimum needed to learn the mapping. However, results were very similar over a range of hidden unit numbers. Unit activations were calculated with the hyperbolic tangent function,

$$a_j^{[t]} = \tanh\left[\gamma\left(I_j^{[t-1]} + \Delta t\left(I_j^{[t]} - I_j^{[t-1]}\right)\right)\right], \tag{1}$$

where γ was input gain, Δt was an integration constant fixed at 0.166, and $I_j^{[t]}$ was the net input at time t. For the feed-forward network, there was no time course of processing, so there was no integration constant, and activations were simply a function of the instantaneous net input. Input gain was fixed at 1 during training, and varied during testing (see following section). The net input to each unit was calculated as the dot product between the activation vector over its sending units, and the weight vector over its incoming connections. The hyperbolic tangent is a sigmoidal function with asymptotes at +1 and −1.

Forward connection weights were initialized to random values in the range ±0.1, and recurrent weights (for the recurrent model) were initialized in the range ±0.5. A larger range was used for recurrent weights to ensure that they had a substantial impact on processing. Weights were learned by gradient descent,

$$\Delta w_{ij} = \eta\left(\partial E/\partial w_{ij}\right), \tag{2}$$

where w_{ij} was the connection weight from unit j to i, η was the learning rate (fixed at 0.001), and

E was cross-entropy error (Rumelhart, Durbin, Golden, & Chauvin, 1995), which was scaled by each item's frequency. Weight changes were made after each time weight derivatives had been accumulated over all 1024 items in the corpus. Weight derivatives were calculated for each item as follows: input units were set to a given item's input pattern, activations were propagated forward through the network, and an error signal was calculated from the difference between actual and target outputs. In the feed-forward model the error signal was then backpropagated to generate the weight derivatives. In the recurrent model, activations were propagated forward for 18 ticks, error was injected on the last 12 ticks, and then error was backpropagated in time.

Weight updates were repeated until every output unit was within 0.1 of its target for every item in the training corpus. This criterion was reached in the feed-forward model after 62,000 passes through the corpus, and in the recurrent model after 56,000 passes.

2.3. Testing procedure

The models were tested by setting the input units to a given input pattern, and recording the output activations. In the recurrent network, activations were recorded on the first tick for which all 12 output nodes were within 0.1 of an asymptote. If a node did not reach this criterion after 18 ticks, the output was judged as incorrect. The criterion for correct performance was having the activations of all 12 output units on the target side of zero. Targets for items in the corpus were set according to each item's output pattern. Targets for the 3072 novel items were set according to each item's input pattern, i.e., the identity mapping.

To dissociate item-based and regularity-based processing, input gain was varied as a single control parameter over the hidden units. The reported levels of input gain were between 0.33 and 3. This range was chosen to show asymptotic performance at the lower and upper ends, i.e., the patterns of behavior did not change substantially beyond this range.

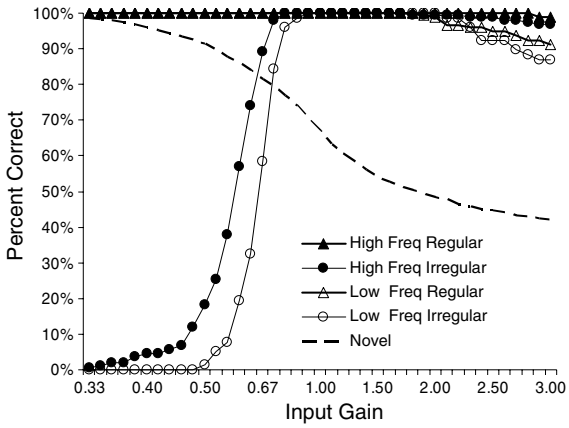

Fig. 2. Percent correct performance of the feed-forward model as a function of input gain and item type.

3. Simulation results

To simplify the presentation of results, known items were divided into categories of high frequency and low frequency. The high frequency category consisted of the 256 most frequent trained items (top quartile), and the low frequency category consisted of the 256 least frequent trained items (bottom quartile). Mean accuracies for the feed-forward model are graphed in Fig. 2 as a function of input gain and item type (high and low frequency regular, high and low frequency irregular, and novel). The same are graphed for the recurrent model in Fig. 3.

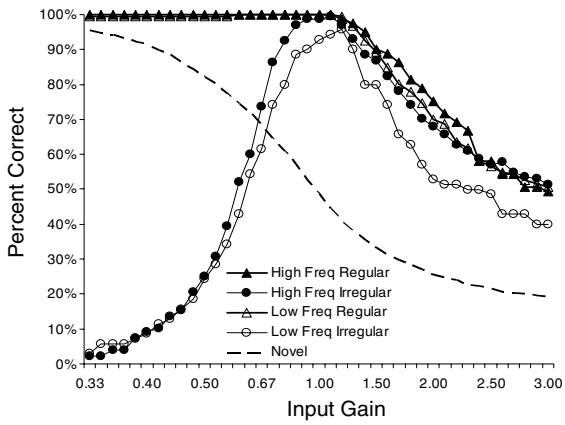

Fig. 3. Percent correct performance of the recurrent model as a function of input gain and item type.

Figs. 2 and 3 show that both models exhibited a clear dissociation in performance on irregular items compared with novel items. At low levels of input gain, generalization of the identity mapping to novel inputs was essentially perfect, as was performance on regular items. By contrast, performance on irregular items dropped to 0%, at which point all inputs were computed as the identity mapping. For irregular items, application of the identity mapping can be considered as a *regularization* error because, for the quasi-regular domain constructed here, the identity mapping is the regular mapping.

At high levels of input gain, performance on known items was better than performance on novel items. At their maximum difference, mean accuracies in the feed-forward model were 97% for known items, and 46% for novel items. The same comparison in the recurrent model was 94% and 35%, respectively. The biggest difference between the models was that performance on known items was near ceiling at all high levels of input gain in the feed-forward model, whereas performance on known items dropped off at sufficiently high levels of input gain in the recurrent model. This difference is explained by the fact that high levels of input gain amplify the effect of nonlinearities in the activation function on processing (see below). Recurrence caused this amplification to increase to the point of distorting the course of processing. Given that the dissociation was maintained despite this distorting effect, we did not see it as informative with respect to the dissociating effect of input gain.

The models' behavior was reminiscent of the double dissociation seen in surface and phonological dyslexia. Table 1 compares the word naming performance of two surface dyslexic patients (MP, Behrmann & Bub, 1992; KT, McCarthy & Warrington, 1986) with each model's performance at low levels of input gain. Input gain was manipulated as a free parameter in each model to best match each patient's pattern of performance. Similarly, a phonological dyslexic patient (WB, Funnell, 1983) can be compared with both types of models processing at increased levels of input gain. Of known words, the patient WB performed 89% correctly, while the feed-forward and recurrent models correctly performed 93% and 44%, respectively. With novel words, the patient WB performed 0% correctly, while the feed-forward and recurrent models correctly performed 42% and 19%, respectively. The purpose of these comparisons was only to draw a relation between the current models and dyslexia; as stated earlier, the models were not intended to simulate dyslexia.

The results reported here show that the manipulation of input gain as a single control parameter, over a single level of representation, caused a double dissociation in both models. Thus, we can conclude that the dynamics produced by recurrent connectivity is not required to give input gain its dissociating effect. While this result is informative, it does not fully elucidate the computational principles by which input gain has its dissociating effect. The following analyses were designed to explicitly show that input gain affects the extent to which processing is componential versus conjunctive, and that it is this property of input gain that is responsible for its dissociating effect.

3.1. Componential versus conjunctive processing

In the context of the current models, componential processing occurs when each input dimension is used independently of all other input

Table 1
Surface dyslexic MP and KT, compared with feed-forward (FF) and recurrent (Re) models at various decreased levels of input gain, on high frequency (HF), low frequency (LF), regular (Reg), irregular (Irr), and novel items

	MP (%)	FF (%)	Re (%)	KT (%)	FF (%)	Re (%)
HF Reg	95	100	100	100	100	100
LF Reg	98	100	100	89	100	99
HF Irr	93	89	86	47	57	25
LF Irr	73	58	74	26	19	24
Novel	96	82	65	100	86	82

dimension in computing an output pattern. That is, each input dimension is treated as an independent component of the input pattern. By contrast, conjunctive processing occurs when conjunctions of input dimension are used to compute outputs. In this section, we first demonstrate that low levels of input gain shifted the models into a componential mode of processing, whereas high levels of input shifted them into a conjunctive mode of processing. We then explain why input gain had this effect, and why it was responsible for the observed dissociation.

The degree of conjunctivity in processing was measured by manipulating the value of a single input dimension, while holding the remaining dimensions constant at a neutral value of 0.5. Thus, the manipulated dimension was responsible for any changes in output activations. As a given input dimension was manipulated, activation values were monitored for the 11 output units that did not correspond to the manipulated input dimension. The distance from baseline for each activation value (i.e., closeness to asymptote) was then calculated, and these distances were averaged across all output dimensions, for manipulations of all 12 input dimensions. This calculation was a measure of the conjunctivity of processing because higher values indicated that input dimensions were having greater effects on non-corresponding output dimensions.

The measure of conjunctivity is plotted in Fig. 4 as a function of input gain and model type. The figure shows, in both models, the conjunctivity of processing increased with higher levels of input gain. In the next two sections, we explain why componential processing at low levels of input gain could not support the processing of irregular items, and why conjunctive processing at the high levels of input gain could not support the processing of novel items.

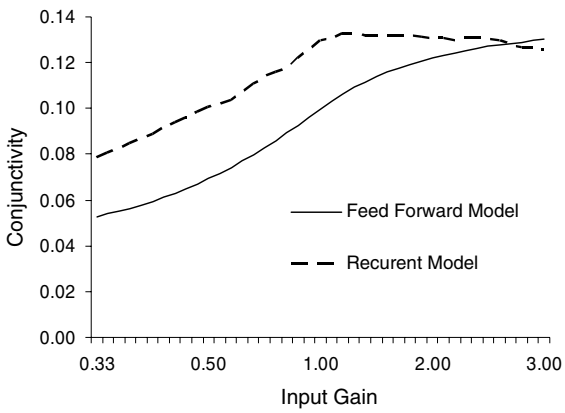

Fig. 4. A measure of conjunctive processing in the feed-forward and recurrent model as a function of input gain.

3.2. Low levels of input gain

Low levels of input gain caused the sigmoidal shape of the hyperbolic tangent function to flatten and become more linear. Linear hidden units can only support a linearly separable mapping between the inputs and outputs. It is also the case that input dimensions are processed as independent components in a linearly separable mapping. In the quasi-regular mapping created for these models, the identity mapping was linearly separable, whereas the exceptions to the identity mapping were not.

Moreover, hidden units operated in their linear range early in training because the positive and negative random initial weights tended to cancel each other out on any given input; therefore, net inputs to the hidden units tended to be small early in training, as if input gain was low (even though input gain remained fixed at 1 throughout training). The upshot of these points is that the linearly separable identity mapping was learned in the linear range of the hidden units early in training, and exceptions to the identity mapping were learned only after hidden unit activations moved closer to their asymptotes. An analysis confirmed this statement: after only 30 epochs of learning, the model applied the identity mapping to all 4096 possible input patterns, and hidden unit activations were 0.35 away from zero on average, i.e., they were mostly operating in their linear range. Therefore, one can infer that low levels of input gain invoked the componential, linearly-separable identity mapping that was learned early in training.

3.3. High levels of input gain

In contrast with low levels of input gain, high levels cause the sigmoidal shape of the activation function to sharpen and more closely mimic a step

function. The consequence of this sharpening is that hidden unit activations are more likely to be near their asymptotes. This point is important because it is the asymptotic behavior of hidden units that enables them to use conjunctions of input values in mapping their inputs onto their outputs (O'Reilly, 2001). The XOR function is the quintessential example in which the conjunction of two input values must be considered in order to produce the correct output, and at least one nonlinear hidden unit is necessary to compute this function in a feed-forward neural network. In the quasi-regular mapping learned by the distributed model, it was the irregularities that engaged the asymptotic behavior of the hidden units in order to process conjunctions of input values. On this logic, one can say that high levels of input gain placed a greater emphasis on conjunctive processing of input values.

Given this functional effect of input gain, why would an emphasis on conjunctive processing cause a selective impairment in the processing of novel input patterns? The answer begins with the fact that the conjunctive processing was used to handle irregularity in the quasi-regular mapping, whereas, it is the regularity in this mapping that provides the basis of correct performance on novel inputs. Therefore, the conjunctions learned for irregular inputs will tend to be incorrectly applied to novel inputs at high levels of input gain.

This logic leads one to ask, why was performance on regular known items intact at high levels of input gain? If conjunctions were learned only to process irregularities, then one would have to conclude that the emphasis on conjunctions at high levels of input gain should interfere with the regular identity mapping for all input patterns, both novel and known. What is missing here is that conjunctions were learned not only to process the irregularities for known items, but regularities as well. This consequence of using conjunctions necessarily followed from the use of distributed representations over the hidden units. When the asymptotic behavior of hidden units was engaged through learning, it affected the processing of all known inputs, both irregular and regular, because every hidden unit contributed to the processing of every input dimension, for every input pattern. Therefore, conjunctions had to be learned for the regularities in the known items, and these conjunctions supported correct performance on known items at high levels of input gain.

4. Conclusions

The current simulations demonstrated how a double dissociation can occur in a non-modular system via the manipulation of a control parameter. Analyses showed that input gain shifted the current models between conjunctive and componential modes of processing. This shift produced a pattern of behavior that was reminiscent of the word naming impairments that characterize phonological and surface dyslexia. Conjuctive modes impaired the processing of novel items, whereas componential modes impaired the processing of irregular items.

The current work is not intended as an explanation of any particular language pathology. It is unclear whether input gain would provide a satisfying account of specific empirical results. For instance, input gain would not appear to handle dissociations in which all regular items, both novel and known, are impaired (Marslen-Wilson & Tyler, 1997, 1998; Ullman, Corkin, Coppola, & Hickok, 1997). However, these simulations make a strong statement against an often utilized interpretation of double dissociations. While a modular system offers a transparent explanation of a double dissociation, the observation of such a phenomenon should not be interpreted as necessitating separable components.

Further, empirical and computational work is necessary to determine whether input gain can provide the best available account for certain cases of dissociations. Some of this work would need to formulate the neural basis of input gain. There are several plausible candidates for the neural mechanism of sensitivity-modulation which is analogues to the function of input gain in the current models. For instance, it is well-established that some neuromodulators can change the sensitivity of neurons to their inputs (see Fellous & Linster (1998)). Impairments in sublexical or lexical

processing might be explained by brain damage that disrupts the function of an appropriate neuromodulator.

Another possible neural basis of input gain is a modulatory system, rather than a neuromodulator. Such a system would serve to modulate the sensitivity of neurons in networks outside of itself, with the purpose of regulating neural systems, or adapting them to changes in task demands. If a modulatory system existed to regulate the balance of componential and conjuctive processing via input gain, then damage to the modulatory system might cause "mode locking" into either componential or conjunctive processing.These ideas are speculative for now, but they demonstrate the plausibility of an input gain account of surface and phonological dyslexia.

The current simulations and the single-route model of word reading (Kello, 2003; Kello & Plaut, 2003) comprise the beginnings of a single-route alternative to dual-route theories of word reading. They address one of the biggest challenges to any single-route theory of word reading, namely, the neuropsychological evidence for separable sublexical and lexical processing routes. However, many challenges remain. How would a single-route alternative be consistent with evidence for separable brain regions correlated with sublexical and lexical processing, to the extent that such evidence exists? Are there any testable differences between explanations of double dissociations using control parameters and those using separable components? Could a large-scale, single-route model account for any of the more detailed findings in the vast literature on word reading? These questions await further research.

References

Behrmann, M., & Bub, D. (1992). Surface dyslexia and dysgraphia: Dual routes, single lexicon. *Cognitive Neuropsychology, 9*, 209–251.

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review, 100*, 589–608.

Coltheart, M., 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.

Fellous, J.-M., & Linster, C. (1998). Computational models of neuromodulation. *Neural Computation, 10*, 771–805.

Funnell, E. (1983). Phonological processes in reading: New evidence from acquired dyslexia. *British Journal of Psychology, 74*, 159–180.

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading, and dyslexia: Insights from connectionist models. *Psychological Review, 163*, 491–528.

Kello, C. T. (2003). The emergence of a double dissociation in the modulation of a single control parameter in a nonlinear dynamical system. *Cortex, 39*, 132–134.

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

Marslen-Wilson, W. D., & Tyler, L. K. (1997). Dissociating types of mental computation. *Nature, 387*(6633), 592–594.

Marslen-Wilson, W. D., & Tyler, L. K. (1998). Rules, representations, and the English past tense. *Trends in Cognitive Sciences, 2*, 428–435.

McCarthy, R., & Warrington, E. K. (1986). Phonological reading: Phenomena and paradoxes. *Cortex, 22*, 359–380.

O'Reilly, R. C. (2001). Generalization in interactive networks: The benefits of inhibitory competition and Hebbian learning. *Neural Computation, 13*, 1199–1241.

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

Plaut, D. C. (1995). Double dissociation without modularity: Evidence from connectionist neuropsychology. *Journal of Clinical and Experimental Neuropsychology, 17*, 291–321.

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.

Plaut, D. C., & Kello, C. T. (1999). The emergence of phonology and from the interplay of speech comprehension and production: A distributed connectionist approach. In B. MacWhinney (Ed.), *The emercenge of language* (pp. 381–415). Maweh, NJ: Erlbaum.

Rumelhart, D. E., Durbin, R., Golden, R., & Chauvin, Y. (1995). Backpropagation: The basic theory. In C. Yves & D. E. Rumelhart (Eds.), *Backpropagation: Theory, architectures, and applications* (pp. 1–34). Hillsdale, NJ: Lawrence Erlbaum.

Ullman, M. T., Corkin, S., Coppola, M., Hickok, G., et al. (1997). A neural dissociation within language: Evidence that the mental dictionary is part of declarative memory, and that grammatical rules are processed by the procedural system. *Journal of Cognitive Neuroscience, 9*, 266–276.

Van Orden, G. C., Pennington, B. F., & Stone, G. O. (2001). What do double dissociations prove? Cognitive Science, *25*, 111–172.