2 Progress in understanding word reading: Data fitting versus theory building

Mark S. Seidenberg
University of Wisconsin, USA
David C. Plaut
Carnegie Mellon University, USA

Computational modelling is a tool that can be used in different ways for different purposes. There are several distinct styles of modelling research in cognitive science and neuroscience, with differing goals, methods, and evaluation criteria. Nowhere is this clearer that in the area of lexical processing in reading, in which computational models have played a prominent role for over 25 years.

In this chapter, we examine two contrasting approaches to computational models of reading, the dual-route approach developed by Max Coltheart and his colleagues, and the parallel distributed processing (PDP) approach developed by ourselves, James L. McClelland, and others. These approaches have spawned a series of implemented models (including Coltheart, Curtis, Atkins, & Haller, 1993; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Harm & Seidenberg, 1999, 2004; Plaut, 1997; Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg & McClelland, 1989, among others). The usual way this research is assessed is by examining individual models with respect to factors such as their fidelity to behavioural data, the breadth of the behavioural phenomena addressed, the limitations of the model, and so on. Our primary goal in this chapter is not to evaluate individual models but rather to examine more basic foundational assumptions of the two approaches. They differ fundamentally with respect to these assumptions, including what models are for, how they are developed, and how they are to be evaluated. Of course, these differences greatly complicate the task of comparing models. Dual-route models are constructed with specific desiderata in mind, some of which are not shared by the other approach, and they necessarily come out "ahead" if those criteria are used in comparing models. The same is true for the PDP models. If all fruits are judged on the basis of what makes a good apple, then apples are necessarily the best fruit. This is unsatisfying if only some of the criteria also apply to oranges.

Our goal is to clarify the differences between the approaches and how they affect the interpretation of specific models. We will argue that the more important consideration at this stage in the development of the field is

between the alternative approaches, not individual models. The questions that end up mattering are ones such as, which approach raises the most interesting questions? Is better able to relate behaviour to its brain basis? Can explain individual variation in reading ability or style? Offers theoretically meaningful links between reading and other aspects of cognition? Our claim is that other questions, such as which model has been applied to the broader range of experimental paradigms and tasks, or how good a fit has been achieved to the results of specific behavioural studies, are currently somewhat less important. The latter questions already presuppose a commitment to a particular approach, but the basis for making such a commitment is what needs to be examined. One set of questions involves assessing which approach is on a trajectory toward answering basic questions; the other set is more focused on keeping score about the intermediate products of the research programmes.

It will not surprise readers to learn that, in our view, the results of this analysis favour the PDP approach. But convincing readers of this conclusion is far less important to us than facilitating a deeper understanding of the nature of the disagreement. There is no doubt that the dual-route approach has served a highly valuable function by orienting researchers to important research questions, and generating testable hypotheses that promoted numerous empirical studies over a multiyear period. Moreover, the competition between the approaches has advanced the understanding of the theoretical and empirical issues considerably. In our assessment, however, the implemented dual-route models—particularly the 2001 attempt to expand the range of phenomena to which the dual-route cascaded (DRC) model was applied—exposed some fundamental limitations of the approach.

DRC is an example of a bottom-up, data-fitting approach to modelling that has a long history in cognitive science. The limitations of this approach have been widely discussed in the literature, dating from at least Newell's famous "Twenty Questions" chapter (1973). The basic problem with the bottom-up approach is that a model can fit specific data patterns without capturing the principles that govern the phenomena at a biological, computational, or behavioural level. "Fitting the data", then, tells us more about the flexibility of a style of modelling than about the questions that motivated the research in the first place. To be clear, the PDP approach is not wholly immune from these problems either, because it also involves detailed comparisons between models and data, and not all aspects of all models are equally well motivated. However, it largely avoids them by being grounded in a set of more completely specified and constrained computational principles. The emphasis in the PDP approach is not on capturing every empirical data point in a single model but rather on providing a framework for addressing issues that will continue to be the focus of attention for the foreseeable future: how the brain achieves the computations that underlie reading; the relationship between reading and other capacities; the bases of differences (across individuals and writing systems) in reading, and how such differences interact

with brain injury; and the causes of developmental dyslexia, understood at genetic, neurophysiological, and behavioural levels.

The dual-route approach

The term "dual-route model" is ambiguous and a source of confusion for many researchers (Coltheart, 2000; Harm & Seidenberg, 2004). The term can refer either to dual (visual versus phonological) mechanisms for accessing the meanings of words from print (which Coltheart terms "DR-M models") or dual (lexical versus sublexical) mechanisms for computing the pronunciations of words (which Coltheart terms "DR-P models"). Although Coltheart (1978) discussed the issue of visual versus phonological processes in reading, most subsequent research focused on the DR-P idea (e.g. Coltheart, Davelaar, Jonasson, & Besner, 1977; Paap & Noel, 1991; Patterson, Marshall, & Coltheart, 1985), and that is our focus here. The dual-route theory was initially developed by means of informal information-processing models (sometimes called "box and arrow" models), of which there were multiple variants (see Coltheart, Sartori, & Job, 1987; Patterson et al., 1985, for examples).

Computational versions of the DR-P model were eventually developed after critiques of the informal modelling approach (e.g. Seidenberg, 1988) and the development of connectionist models of word reading (e.g. McClelland & Rumelhart, 1981; Seidenberg & McClelland, 1989).

Coltheart et al. (2001) presented considerable background concerning the origins of their approach and its fundamental assumptions. They link their models to nineteenth-century "diagram makers" such as Lichtheim (1885), and emphasize the continuity between the informal (e.g. Coltheart et al., 1977) and computational versions of the dual-route model. Principal features of the approach include the commitment to a version of the modularity hypothesis (Fodor, 1983; Coltheart, 1999), to theorizing pitched at the level of the "functional architecture" of the cognitive system (Shallice, 1988), and to identifying the modules of the functional architecture primarily through studying brain-injured patients.

Several other important components of the approach should also be noted. First, Coltheart et al. (2001) emphasize the data-driven character of their modeling, endorsing Grainger and Jacobs' (1998, p. 24) view that "in developing algorithmic models of cognitive phenomena, the major source of constraint is currently provided by human behavioral data". Second, they view models as cumulative: each model improves upon the previous one by adding to the phenomena covered by a previous version. Thus, the 2001 version of the DRC is said to account for the same facts as the 1999 version but also many others. The models are described as "nested" with respect to data coverage. Again they quote Grainger and Jacobs (1998): "[I]n other sciences it is standard practice that a new model accounts for the crucial effects accounted for by the previous generation of the same or competing models." Finally,

Coltheart and colleagues emphasize fidelity to behavioural data as the principal criterion for evaluating models (see also *Rastle & Coltheart*, this volume). Models are designed to simulate data patterns, and a model is valid unless disconfirmed by making an incorrect prediction. So, for example, in justifying their use of an interactive-activation (IA) model as a component of DRC, the authors note that the McClelland and Rumelhart IA model had not been "refuted" by any behavioural data; thus, there was no empirical reason to abandon it.

Having described the rationale behind their approach and motivated major parts of their model's architecture, Coltheart et al. (2001) turned to applying the model to a large number of phenomena. The most noteworthy feature of the research was the application of a single model to more than 20 empirical phenomena involving word and nonword reading. The breadth of the data coverage led Coltheart and colleagues to conclude that "the DRC model is the most successful of the existing computational models of reading" (p. 204).

Comments on the approach

As already noted, the technique of fitting models to data has a long history in psychology. Other prominent examples from the not-too-distant past include Sternberg's (1969) serial search model; Smith, Shoben, and Rips' (1974) model of semantic categorization; Collins and Loftus' (1975) spreading activation model of semantic memory; Clark's (1969) deductive reasoning model; Ratcliff's (1978) diffusion model of memory retrieval; Massaro's (1989) fuzzy logical model of perception; and so on. The methodology in this research was roughly as follows: behavioural data were collected in the service of addressing a general question such as "how do people retrieve information from memory?" or "how are categories organized in memory?" The data derived from a small number of laboratory tasks such as probe verification ("is this stimulus a member of a prespecified target set?"), statement verification ("is a canary a bird?"), and probe recognition ("is this stimulus new or old?"). The researcher then developed a model that rationalized various facets of the data, such as why some statements are more difficult to verify than others. The models were usually presented informally (in words or figures), although some models were implemented as computer simulations for additional precision and clarity. The models were mainly tools for developing explicit accounts of the types of procedures and knowledge representations that gave rise to overt performance.

These models were developed out of a vocabulary of representations and processes loosely drawn from flowcharting. Thus, there would be data structures (such as propositions) and operations upon them, such as encoding, comparing, testing conditions, decision making, and so on. The inventory of elements out of which a model could be constructed was large, providing considerable flexibility and descriptive power. A model could then be

developed which fit some interesting empirical phenomenon (For example, why it is easier to decide that a canary is a bird than an animal).

The DRC models were developed in a similar fashion. Over a multiyear period, researchers had collected a large body of data in the service of addressing basic questions such as how people read words aloud (naming) or decide whether a string of letters is a word (lexical decision). Informal models were then constructed to fit these data. The models were built out of constructs, such as rules (for pronouncing letter strings), lexical nodes, activation levels, spreading activation, and so on, that were carried over from earlier accounts of word reading. These concepts were employed because they closely matched intuitions about reading and language, and because they were they were available as part of the information-processing vocabulary of the era. Eventually, the models were implemented as computer simulations that reproduced behavioural effects, such demonstrations being taken as evidence that the model was correct, especially when a single model could be applied to a large number of findings derived from several tasks.

Coltheart et al. (2001) termed this approach "old cognitivism" (in contrast to the more recent PDP paradigm). Although they correctly situated their research within this older tradition, they did not discuss long-standing critiques of it. The basic problem is that the strategy of accounting for the most data possible is not itself sufficiently powerful to converge on a satisfactory theory in cognition. Rather, what happens is this: The researcher has a certain favoured vocabulary of elements out of which to construct a model that accounts for target behavioural data. There is also considerable freedom to introduce new types of representations, processing mechanisms, assumptions about the time course of processing, parameter settings, and so on (see, for example, the discussion in Seidenberg, 1988, of how boxes are added to such models). Considerable resources are therefore available for model construction. The elements of the model are configured in response to empirical data; that is the essence of the methodology. This procedure does not guarantee, or even promote, converging on theoretical generalizations that explain the phenomena.

This assertion seems rather harsh; how can it be evaluated? One way is to ask of any given model whether it accounts for data other than those that were used in constructing it. If a model instantiates relevant general principles, it should accommodate other data of a similar kind. If the model merely fits selected target data, there is no guarantee that it will generalize; and if it does not, this is a telltale sign that the model has failed to capture the general principles that underlie the behaviour. In our view, the DRC model—particularly the 2001 version—is subject to this criticism.

The DRC 2001 model was configured to fit the results of studies illustrating a variety of phenomena concerning word and nonword reading. For each phenomenon (such as regularity effects, frequency by regularity interaction, consistency effects, nonword naming, pseudohomophone effects, and so on), the authors fit the results of a key experiment or two demonstrating the

effect. For example, the frequency by regularity experiment was one by Paap and Noel (1991), the consistency experiment was one by Jared (1997), and the pseudohomophone experiments were from McCann and Besner (1987) and Taft and Russell (1992). In each case, of course, one might choose other experiments addressing the same phenomenon. For example, the frequency-by-regularity interaction can also be assessed with studies such as Jared (1997); Seidenberg (1985); Seidenberg, Waters, Barnes, and Tanenhaus (1984); Taraban and McClelland (1987); Waters and Seidenberg (1985); and others. Consistency effects can be assessed with respect to studies such as Cortese and Simpson (2000); Jared (2002); Jared, McRae, and Seidenberg (1990); and Seidenberg et al. (1984). If one examines the results across a set of studies on one issue, it often turns out that the behavioural effect in question consistently replicates across studies, but not in DRC. In other words, DRC's fit to the data is rather better for the study presented in Coltheart et al. (2001) than for other exemplars.[1]

To illustrate, a number of studies have reported that words with inconsistent spelling-sound correspondences yield longer naming latencies than words with consistent correspondences (e.g. Cortese & Simpson, 2000; Glushko, 1979; Jared, 1997, 2002; Seidenberg et al., 1984; Waters & Seidenberg, 1985). A word such as GAVE is inconsistent because the pattern–AVE is pronounced differently in HAVE. Inconsistent words are theoretically important because they are rule governed and therefore should pattern with words such as MUST, which have no close inconsistent neighbours. The weights in a PDP model reflect the net effects of exposure to a large corpus of words. Other factors being equal, inconsistent words such as GAVE should take longer to read aloud than a more consistent words such as MUST, even though both are "rule governed". This effect was first reported by Glushko (1979); subsequent studies showed that it is larger for lower frequency words and for younger readers. Consistency effects therefore provide evidence to support a model which represents degrees of consistency in the mapping from spelling to sound, rather than rule-governed forms and exceptions (Seidenberg & McClelland, 1989). However, Coltheart et al. (2001) suggested that consistency effects arise from other, confounding factors, and showed that DRC simulated the results of a study by Jared (1997) (Figure 2.1). Thus, consistency effects seemed to be less of a problem for the dual route approach than some had thought.[2]

Consistency effects have been reported in several other studies; how does DRC fare with respect to these findings? Consider Glushko's (1970) original study comparing regular words (e.g. MUST), exception words (e.g. HAVE), and inconsistent words (e.g. MINT). As Figure 2.2 indicates, the study yielded a regularity effect (regular words faster than irregular) but also a consistency effect (inconsistent words longer than regular). The DRC model, tested on the same words, reproduces the regularity effect, but not the consistency effect (Figure 2.3). Jared (2002) manipulated both regularity (rule-governed words versus exceptions) and consistency (consistent versus

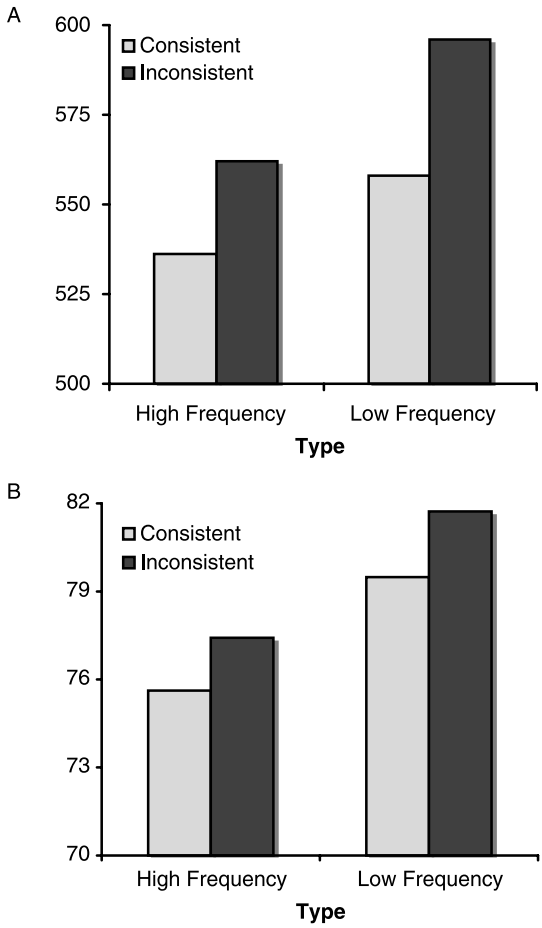

Figure 2.1 (A) Results of Jared's (1997) study of consistency effects. (B) Dual-route cascaded simulation of the same study.

inconsistent spelling patterns) in a factorial design. The behavioural results and DRC's performance on the same items are presented in Figure 2.3. The behavioural data yielded a consistency effect (consistent < inconsistent) but no effect of regularity. The simulation yielded a regularity effect (regular < exception) but no effect of consistency. There are similar differences between the behavioural and simulation results in other studies of consistency effects.

Thus, the DRC model does not capture the consistency effects observed across different studies in different laboratories. The methodology used in developing DRC results in overfitting: The model is closely tailored to the results of specific experiments and is less successful when assessed against other studies. Would one really want to conclude that a model captures a phenomenon if it adequately reproduces the results of one study, but not

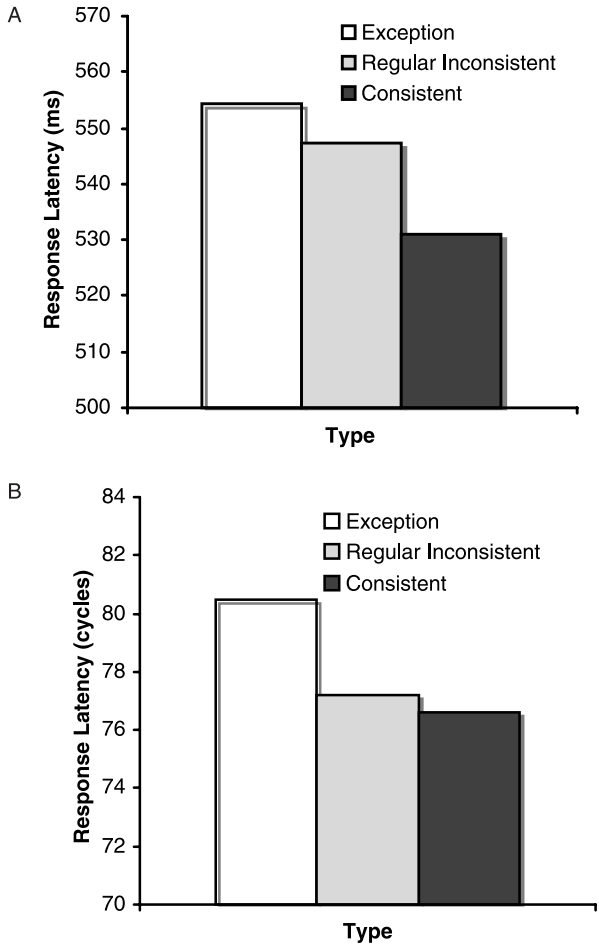

Figure 2.2 (A) Glushko's (1979) study of regular (rule-governed), regular inconsistent, and exception word naming. Latencies for regular inconsistent words fall between regular words and exceptions. (B) Dual-route cascaded simulation using the same words. Regular inconsistent words do not differ from regulars.

others that yielded the same pattern of results (e.g. using different stimulus items)? One could just as well conclude that the model failed to simulate the phenomenon.

One might further hold that a model needs to explain phenomena in a principled way. That is, the explanation for the phenomena should derive from biological, computational, or behavioural principles that have some independent justification and thus are not merely introduced in response to the data at hand. Judged by these additional criteria, DRC does not fare well either. DRC's core commitments were introduced as ways of rationalizing

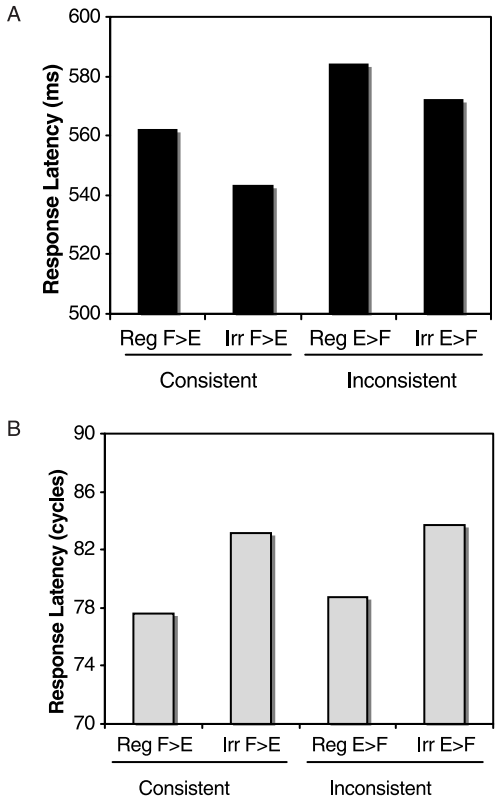

Figure 2.3 (A) Results of Jared's (2002) study of regularity and consistency effects. F = friends, E = enemies. Words with more friends than enemies are regular; words with more enemies than friends are exceptions. The results indicate a reliable effect of consistency and no effect of regularity. (B) Dual-route cascaded simulation of this study. The results show a reliable effect of regularity but not consistency.

data about reading, not because they were independently motivated by other concerns. They were largely inherited from the informal models of the pre-computational era that relied heavily on intuitions about reading: for example, the "lexical entries" of the 1970s became the "localist representations of words" in the DRC models.

We have noted that DRC-type models are built out of a large inventory of modelling elements. There is little constraint on what can be included in a model to allow it to fit data successfully. It might therefore seem that the DRC model merely needs some additional modification in order to fix the generalization problem identified above. For example, interactive-activation models (one component of DRC) have a large number of parameters governing the flow of activation between and within units. (Coltheart et al., 2001, Table 1, p. 218, list 31 such parameters, which do not include seven additional

parameters introduced to model lexical decision.) Each of these parameters can take on a large number of values. Hence, the space of IA models defined by these parameters is huge. Perhaps fixing the model is just a matter of searching this space more systematically to find a set of parameters that yields good fits to a broader range of behavioural results, or of adding additional components to the model. Indeed, it is common lore that complex models with large numbers of parameters can "fit any data"; all that is required is sufficient skill and stamina to find the right combination.

It turns out that DRC-type models do not actually behave this way. Rather, developing a model to fit a set of data has the opposite effect: it makes it more difficult to modify the model to accommodate additional data. The results described by Coltheart et al. (2001) are closely tied to the parameter settings that were used. We have not explored all possible combinations of parameter settings, but it is clear from experimentation with the model that it is very difficult to find a different parameter set that yields better results than the ones reported. Although better fits might be achieved, one assumes that if the authors had been able to find such a set, they would have reported it. In fact, the model's performance is brittle insofar as it does not perform as well with many other parameter settings.

Further complications arise from altering parameters to account for additional phenomena. For example, in Coltheart et al. (2001), the standard parameters failed to exhibit the facilitatory effect of neighbourhood size on word-naming latencies observed empirically by Andrews (1989, 1992) and by Sears, Hino, and Lupker (1995). In order to get the DRC to exhibit this effect, Coltheart and colleagues eliminated lateral inhibition within both the orthographic and phonological lexicons, and reduced feed-forward letter-to-word inhibition. Even so, while this yielded a main effect of neighbourhood size, it still failed to produce the interaction of neighbourhood size and word frequency as observed by Andrews (1992) and Sears et al. (1995). Moreover, these new parameter values create problems with respect to other phenomena. For example, the new parameter regime introduced to produce neighbourhood effects compromises the ability of the model to exhibit consistency effects. Whereas the original parameters produced a weak but reliable effect of consistency for Jared's (1997, experiment 1) stimuli (and no errors), the new parameters produce relatively high error rates (10–16% across conditions) and no reliable consistency effect.

This behaviour is characteristic of DRC. Changing parameters may allow the model to fit better the results of an individual experiment. However, it is then essential to examine the effects of these parameter changes on *other* aspects of the model's behaviour, which Coltheart et al. (2001) did not report. In practice, changing parameters to fit specific results has the effect of moving the behavioural mismatch somewhere else. One leak is patched but others spring up elsewhere because the underlying problem has not been addressed.

This analysis is further borne out by developments since the publication of the 2001 version of DRC. Coltheart and colleagues have presented additional

data and arguments favouring the DRC model, primarily focused on atypical cases of developmental or acquired dyslexia with patterns of impaired reading that are said to contradict one or another aspect of the PDP theory (e.g. Blazely, Coltheart, & Casey, 2005; Coltheart, in press). The interpretation of such case studies is highly controversial, particularly double dissociations (see Dunn & Kirsner, 2003; Juola & Plunkett, 2000; Plaut, 1995; Van Orden, Pennington, & Stone, 2001). The more important point, however, is that Coltheart and colleagues' analyses of these cases are not tied to their implemented model. They wish to argue that certain patterns of behavioural impairment are consistent with their model (and not the PDP approach); however, these arguments are not coupled to simulations of the cases in question.

On our analysis, the successes of the 2001 version of DRC are so tightly bound to particular parameter values and other implementation-specific features that it should be difficult to extend it to other phenomena, including these case studies. The net result is that the most recent arguments in favour of the dual-route approach have the same form as in the precomputational modelling era. We are back to arguments of the 1980s from double dissociations and patterns of "selective" impairment. It is almost as though the models (and the insights about the interpretation of such cases that emerged from them) had never occurred.

The PDP approach

All cognitive processes, including word reading, are ultimately implemented in terms of changes in patterns of neural activity in the brain. Traditional cognitive modelling, of the sort exemplified by the DRC model, assumes that this fact places little if any constraint on the nature of the computations performed by the brain, thereby licensing the introduction of any types of cognitive mechanisms thought capable of fitting behavioural data.

The PDP approach starts from a very different assumption—namely, that the nature of cognitive processing is shaped and constrained in fundamental ways by properties of the underlying neural substrate. The goal is to formulate a set of computational principles that capture how neural activity gives rise to cognition. In pursuing this goal, the mechanisms that are available for modelling are extremely limited and must ultimately be grounded in emerging insights from neuroscience. Note, however, that neuroscientific investigation per se does not directly identify the relevant principles, because not all properties of the brain are equally relevant to understanding cognition. Thus, the PDP approach puts forth *hypotheses* about which properties are central and which are peripheral. These hypotheses are supported by computational demonstrations that models embodying the proposed principles are, in fact, consistent with the relevant behavioural (and neuroscientific) data. Similarly, the limitations of a model can suggest ways in which existing principles need to be modified, supplemented, or replaced. In this way, an individual model

need not be interpreted as a proposed "solution" that is somehow correct or incorrect on the basis of its fit to data. Rather, modelling as an enterprise can be a means of exploring the validity and implications of a set of hypotheses for how cognitive processes are implemented in the brain. It is the underlying hypotheses, rather than the models per se, which collectively constitute a theory of a given domain.

The principles at the core of PDP modelling are well known: cooperative and competitive interactions among simple, neuron-like processing units; different types of information represented by distributed patterns of activity over different groups of units, with similarity reflected by pattern overlap; knowledge encoded as weights on connections between units; and learning as gradual adjustment of connection weights based on the statistical structure among inputs and outputs, often via internal ("hidden") representations. Although this list is certainly not exhaustive—for instance, it makes no mention of intrinsic variability or recurrent connectivity—it conveys the essence of the approach.

Of particular note is the emphasis on learning. Although it is sometimes possible to hand-specify connection weights for small networks that employ localist representations (that is, one unit per entity, as in the lexical route of the DRC model), this quickly becomes infeasible for networks with distributed representations and tens, if not hundreds, of thousands of connections. More to the point, hand-specifying representations, even if possible, would undermine one of the most interesting aspects of the approach—the ability to use a general learning procedure to acquire knowledge that gives rise to effective representations and processes. This property—lacking in many traditional cognitive models including the DRC—is critical for two reasons. The first is that it allows the approach to make direct contact with developmental data on the acquisition of cognitive skills, which can be essential to improving pedagogical practice and elucidating mature mechanisms in adults (e.g. Harm & Seidenberg, 1999; Thomas & Karmiloff-Smith, 2003). The second reason is that the nature of the representations and processes developed through learning, when analysed and understood thoroughly, can give rise to novel and interesting hypotheses about cognitive and neural representations in the brain (Plaut & McClelland, 2000).

A major attraction of the PDP approach is that the same computational principles are brought to bear not only across development, skilled performance, and breakdown after brain damage, but also across the full range of perceptual, cognitive, and motor domains. One consequence is that the mechanisms available for modelling are not introduced solely in response to the data from a particular domain, but are constrained to be broadly consistent with applications in other domains. Of course, developing any specific model involves combining the domain-general principles with domain-specific assumptions about the tasks to be performed and the way in which relevant inputs and outputs are represented. For instance, a simulation of reading aloud must, at the very least, specify a corpus of text for reading acquisition

as well as representations for visual or orthographic input and phonological or articulatory output. Ideally, these domain-specific assumptions are supported by independent evidence, but they can also be taken as additional hypotheses that are subject to evaluation and revision. Thus, it is often informative to develop and examine a range of models that vary in particular ways, in order to clarify what specific properties lead to its ability (or inability) to account for behavioural findings, and the extent to which these derive from domain-specific or domain-general assumptions. In this way, PDP models are like experiments—a means of testing hypotheses by exploring the consequences of particular factors that influence performance.

Finally, it is important to emphasize that, like all models, PDP models are approximations of the representations and processes that underlie human cognition. These approximations are of two kinds. The first concerns the nature of the inputs and outputs of a model. No model implements the entire processing stream from sensory receptors to muscle contractions. Models must incorporate assumptions about the form of input representations generated by earlier, upstream processes, and about how the model's output representations are used by later, downstream processes to produce behaviour. Most word-reading models assume static orthographic input representations and static phonological output representations, but, clearly, these assumptions ignore the complex, temporal nature of early visual processes (including eye movements) (e.g. Rayner, 1998; Reichle, Pollatsek, & Rayner, 1998) and later articulatory processes (e.g. Browman & Goldstein, 1990; Plaut & Kello, 1999). Although omitting these processes may limit the ability of implemented models to address some order and length effects (see *Rastle & Coltheart*, this volume), such effects do not pose fundamental challenges to the more general framework.

The second kind of approximation concerns simplifications within the implementation itself. Not every principle plays a critical role in explaining every phenomenon. Some models make simplifications for reasons of computational efficiency. For example, a feed-forward network is far less computationally demanding to simulate than a fully recurrent version and provides an adequate approximation in some contexts, even though it lacks the interactivity that plays a major role in many PDP accounts. Other simplifications are made for expository purposes. For instance, *Rastle and Coltheart* (this volume) portray Plaut et al.'s (1996) use of localist units for graphemes and phonemes as inconsistent with the principle of distributed representation. However, the simulations were focused on the implications of distributed representation at the level of words and nonwords, and coding graphemes and phonemes as patterns of activity, although straightforward computationally, would have obscured the importance of condensing spelling-sound regularities.[3] Finally, some simplifications provide a means of hypothesis testing. A given model may incorporate only a subset of the relevant principles in order to evaluate whether that subset is sufficient. In this way, modelling is no different than empirical work: no experiment on word recognition examines

all factors that influence processing; rather, studies are designed to provide clear evidence about individual factors and their interactions.

Comments on the approach

The PDP approach to cognitive modelling in general, and word reading in particular, carries with it a number of implications that are worth spelling out in detail. The first and most obvious is that the development of a model is subject to several constraints, only one of which is fitting specific behavioural findings. The result is that, in the short run, specific PDP models may not match particular empirical findings or account for as much variance in empirical data as approaches for which data fitting is the primary goal.

In fact, it would be difficult for PDP modelling to do as well even if data fitting were a goal of the approach. The reason relates to the notion of a *parameter*. In the DRC, a parameter (e.g. letter-to-word inhibition) is a theoretically unconstrained numeric value that can be manipulated independently of other parameters and whose consequences for performance can be evaluated relatively directly. As a result, it is possible to "search" parameter space by running the model repeatedly until a set of parameter values is found that optimizes the fit between model and human performance. The process becomes more computationally demanding when the number of parameters is large and a broad range of data must be fit, but the basic character of the process is unchanged.

The situation is very different for PDP modelling for a number of reasons. First, many of the numeric values within a simulation—most obviously, the connection weights—are not under the direct control of the modeller but are derived algorithmically, and thus do not constitute parameters in any real sense. Second, some of the values that can be directly manipulated, such as those that govern learning and some aspects of network architecture, turn out to have relatively little impact on the nature of what is learned, other than to speed up or slow down the process (as long as the values are within reasonable ranges). Other properties of the models that vary, such as the numbers of units or layers, have broad consequences that are interpretable with respect to behaviour and thus are theoretically relevant.

There are, of course, many aspects of the design of a simulation that ultimately affect its match to behavioural data, but the degree to which these can be optimized is severely constrained by the reliance on learning. Training a large PDP simulation on a complex, realistic task can take days if not weeks on modern-day computing hardware. At this time scale, it is simply infeasible to rerun the simulation a large number of times in order to "search" for parameter values that optimize its fit to data. In practice, small-scale simulations are run to refine general aspects of a simulation, and then multiple runs of the early stages of learning by the full-scale simulation are carried out in order to improve the ability of the model to learn the training corpus. Note, however, that what the model is directly trained to do (generate correct

pronunciations of words) is only a small aspect of how it is evaluated empirically. Rather, models are evaluated largely against incidental consequences of learning, such as the relative rates of acquisition of different types of stimuli, differences in the processing time required to generate word pronunciations, accuracy and error types in pronouncing novel stimuli, or performance after different types of damage. The ways in which design decisions affect these aspects of performance are generally complex and often difficult to anticipate, further limiting the modeller's ability to fit data directly.

None of these comments should be taken to imply that the relationship between model and human performance is not important to the development of PDP models. To the contrary, success at accounting for empirical findings remains the primary means by which a model is evaluated. What must be clarified, however, is what exactly should be taken as the findings against which models should be judged, and what kinds of comparisons between model and human data are informative for understanding the cognitive and neural processes underlying reading.

It should be clear from our earlier discussion of overfitting by the DRC that we consider it unwise to place undue theoretical weight on the quantitative fit of a model to data from any individual empirical study. The results of a single study include large amounts of variance due to a variety of factors other than those under investigation. Even attempts at direct replication that vary only subjects or items often succeed only partially, or reproduce statistical effects but not the quantitative relationship among condition means. Robust behavioural findings that have been replicated across a number of studies provide a more appropriate basis for evaluating models.

Although testing a model with the same stimuli as used in a behavioural study provides a powerful means of comparing the two, care must be taken not to overinterpret the quantitative fit without clear measures of the reliability and precision of the findings themselves. This point is all the more important when considering item-level correlations between model performance and behavioural data from large corpora (e.g. Balota, Cortese, & Sergent-Marshall, 2004; Spieler & Balota, 1997). Even setting aside idiosyncratic aspects of experimental set-up and subject population, the factors that give rise to the most systematic item variance may not be those that are the most important to understand theoretically (Seidenberg & Plaut, 1998). For example, it is well known that word frequency has a strong impact on naming latencies. In PDP models, the influence of frequency on performance is indirect: it determines how often a word is presented during training, which in turn determines how strongly the word's spelling-sound correspondence influences weight values. Although knowledge of all words is superimposed within the same set of weights, the impact that individual words had during training nonetheless leads to faster and more accurate processing. The DRC model lacks a procedure for learning from word presentations, and so (log) frequency is directly stipulated in terms of the resting activation of word units, which, in turn, directly influence processing time. It is thus entirely

unsurprising that the DRC model captures more frequency-related variance than PDP models (Coltheart et al., 2001), but in doing so, it provides no insight into the basis for such effects in man. Conversely, sometimes, factors that account for little item-level variance within a large corpus can nonetheless lead to significant theoretical insight. For instance, the distinction between regularity (whether a word's pronunciation adheres to grapheme-phoneme correspondence rules) and consistency (the degree to which a word's pronunciation agrees with those of similarly spelled words) plays a major role in distinguishing dual-route from PDP approaches, but the amount of response time (RT) variance attributable to either factor is tiny on either account.

The argument for the importance of not overinterpreting individual behavioural experiments applies with equal force to the interpretation of single-case studies in cognitive neuropsychology and their implications for models. Such cases appear to be highly informative because they may produce behavioural effects that are larger and more dramatic than the corresponding effects in the normal population (For example, surface dyslexic patients make errors in pronouncing exception words that normal subjects are merely slower at) or, in some instances, they exhibit effects that are not observed in normals (such as semantic errors in single-word reading by deep dyslexic patients). The standard assumption in interpreting such cases is that all individuals share a common cognitive architecture, so that any observed deviation from control subjects is due solely to the effects of the lesion. This stance is, of course, belied by well-established individual differences in cognitive performance within the normal population in virtually every cognitive domain (see chapters in Boyle & Saklofske, 2004, for overviews). Of immediate relevance is individual variation in tasks such as word and nonword naming (e.g. Andrews & Scarratt, 1998). Zevin and Seidenberg (2006) illustrate how such individual differences can be explained in connectionist networks. The same model was run different times, corresponding to different subjects. Each model was exposed to somewhat different training examples, corresponding to small differences in reading experience. The different models produced different pronunciations for nonwords such as MOUP and WALF, as observed in people (Andrews & Scarratt, 1998; Seidenberg et al., 1994). Hence, the standard assumption underlying the interpretation of the naming performance of brain-injured patients—that there are no relevant premorbid individual differences—is contradicted by the Zevin and Seidenberg (2006) data.

The problem of interpreting single cases is exacerbated by the tendency of neuropsychologists to study (or report) only extreme cases of performance rather than the full distribution of observed cases. Moreover, the alternative of carrying out group analyses of large patient groups is also problematic due to individual differences in lesion characteristics and in the distribution of cognitive processes in the brain (Caramazza, 1986). Perhaps the best compromise is offered by a case series approach, in which behavioural effects are replicated across a series of patients with common aetiology (e.g. Lambon-Ralph, Patterson, & Graham, 2003; Patterson et al., 2006; Rogers,

Lambon-Ralph, Hodges, & Patterson, 2004; Woollams, Lambon-Ralph, Hodges, & Patterson, 2005). On this perspective, atypical findings from single-case studies (e.g. Blazely et al., 2005; Derouesne & Beauvois, 1985) must be viewed with scepticism until replicated in the context of studies that assess a broader range of impaired performance, and theories or models that explain how the full range of cases arises.

Conclusions

The dual-route and PDP approaches to understanding word reading are both supported by explicit computational simulations, but the role that these simulations play in theory development in the two cases is strikingly different. The DRC model of Coltheart et al. (2001) continues the long tradition of a bottom-up, data-driven approach to modelling: A model is designed to account for specific behavioural findings, and its match to those findings is the sole basis for evaluating it. These models aspire to what Chomsky (1965) called "descriptive adequacy". The PDP approach is different. The models are only a means to an end. The goal is a theory that explains behaviour (such as reading) and its brain bases. The models are a tool for developing and exploring the implications of a set of hypotheses concerning the neural basis of cognitive processing. Models are judged not only with respect to their ability to account for robust findings in a particular domain but also with respect to considerations that extend well beyond any single domain. These include the extent to which the same underlying computational principles apply across domains, the extent to which these principles can unify phenomena previously thought to be governed by different principles, the ability of the models to explain how behaviour might arise from a neurophysiological substrate, and so on. The models (and the theories they imperfectly instantiate) aspire to what Chomsky termed "explanatory adequacy". The deeper explanatory force derives from the fact that the architecture, learning, and processing mechanisms are independently motivated (as by facts about the brain) rather than introduced in response to particular phenomena.[4]

The data-fitting approach appears to be better suited to capturing the results of individual studies, because that is the major goal of the approach. DRC thus seems satisfying because it accords with the intuition that accounting for a broad range of behavioural phenomena is always a good thing. Examining DRC more closely, however, suggests two fundamental problems with this strategy. First, the extent to which a model developed in this manner actually fits the data is questionable; fitting the results of one study but not others out of a series is problematic, as are parameter changes that fix one problem but create others. The long list of phenomena that DRC is said to account for needs to be assessed in light of these considerations. The second problem is that the short-term strategy of fitting models to data may in fact contravene the longer-term goal of uncovering fundamental principles. To use an economic metaphor, moves that maximize short-term profits (that is,

fitting more data) may conflict with achieving longer-term economic growth (addressing additional phenomena) and prosperity (converging on the correct theory).

Whereas the DRC approach is data driven, the PDP approach is more theory driven because the models derive from a set of principles concerning neural computation and behaviour. The models are responsive to data insofar as they need to capture patterns that reflect basic characteristic of people's behaviour, particularly with regard to phenomena, such as consistency effects, that distinguish between theories, given their current states of development. The primary goal is not to implement the model that fits the most possible data; rather, it is to use evidence provided by the model, in conjunction with other evidence (such as about brain organization or neurophysiology, or about other types of behaviour) to converge on the correct theory of the phenomena.

The PDP approach to modelling is frustrating to some because there is no single simulation that constitutes *the* model of the domain. The models seem like a moving target: SM89 was interesting but ultimately limited by its phonological representation; PMSP96 largely fixed the phonological problem but introduced the idea that the orth:sem:phon pathway also contributes to pronunciation, something SM89 had not considered. Harm and Seidenberg (1999) used yet another phonological representation and focused on developmental phenomena; Harm and Seidenberg (2004) implemented both orth:sem and orth:phon:sem parts of the triangle but focused on data concerning activation of meaning rather than pronunciation, etc. Each model shares something with all of the others, namely the computational principles discussed above, but each model differs as well. Where, then, is the integrative model that puts the pieces all together?

The answer is, there is none and there is not likely to be one. The concept of a complete, integrative model is a non sequitur, given the nature of the modelling methodology—particularly the need to limit the scope of a model in order (a) to gain interpretable insights from it and (b) to complete a modelling project before the modeller loses interest or dies. The goal of the enterprise, as in the rest of science, is the development of a general theory that *abstracts away from* details of the phenomena to reveal general, fundamental principles (Putnam, 1973). Each model serves to explore a part of this theory-in-progress.

The proponents of DRC view these issues differently. In addition to the data-driven tenet, they emphasize that modelling should be "cumulative", with each successive model addressing the same phenomena as previous models, and new ones as well. Thus each model is a superset of the preceding ones. Here two questions arise. The narrow one is whether the dual-route framework is actually cumulative in the described sense. The second is whether *any* science proceeds in this manner. At first glance, the 2001 version of the DRC appears to be an extension of the 1993 version. In actuality, a variety of changes were made to parameter values and processing assumptions that

render the relationship between the models far more complicated than the "nested" idea suggests. Certainly the two models share core assumptions about the nature of lexical versus sublexical processing, but they differ in the details of how these are implemented. Nor are the models strictly cumulative with respect to the empirical phenomena addressed: for example, the DRC models have used versions of the McClelland and Rumelhart (1981) interactive-activation model as the "lexical" component, and Coltheart et al. (2001) emphasize this genealogy. However, they did not develop DRC by first showing that it accounted for the same phenomena as the IA model (concerning, for example, word superiority effects) and then go on to show that it accounted for others as well.

This situation is not all that different from the relationship among successive versions of the triangle model (Harm & Seidenberg, 1999, 2004; Plaut et al., 1996; Seidenberg & McClelland, 1989): they share a common set of assumptions concerning the architecture of the reading system and basic learning mechanisms, and differ in specific (but important) implementational details. Like DRC, the models address different but overlapping sets of behavioural phenomena and thus are not strictly cumulative.

Whether science proceeds in the cumulative, nested manner described by Grainger and Jacobs (1998) and endorsed by Coltheart et al. (2001) has been strongly questioned, in different ways, by Kuhn (1962) Feyerabend (1975), Lakatos (1970), and many others. Even within the narrower confines of cognitive psychology, it is hard to point to a single example of a series of models that were developed in this strictly cumulative manner. By contrast, it is easy to identify cases in which researchers have developed a series of models that share common principles but differ in detail and address overlapping but nonidentical phenomena (e.g. Roger Ratcliff's diffusion models and John Anderson's ACT models). In short, science does not conform to an incremental, cumulative pattern, nor has the development of the dual-route model.

Finally, *Rastle and Coltheart* (this volume) disparage the exploration of general principles characteristic of the PDP approach as a matter of "faith" and instead endorse modelling on the basis of "inference from evidence". The first of these assertions reflects confusion about our models, and the second confusion about their own. As we noted above, the "principles" that underlie the PDP approach are hypotheses about behaviour and its brain bases. These principles could be proved wrong (via normal scientific methods), in which case they would (and inevitably will) be modified. Whether the principles are correct is thus a question of fact, not an assertion of faith. The second assertion—that models should be developed wholly in response to data—does not accurately characterize the development of scientific theories in general or DRC in particular. Unless science is construed as random fact gathering in the manner of Francis Bacon, empirical studies are invariably conducted in a theoretical context. How did reading researchers decide that it was important to look at words with regular versus irregular pronunciations?

Words do not come labelled as "regular" or "irregular"; the distinction already assumes a nascent theory of the relationships between spelling and sound. In practice, science involves working back and forth between theory and data, with theories guiding what data to gather and the data feeding back on development of the theory. What Rastle and Coltheart apparently mean is that it is valid to introduce an element into a theory (or model) in response to specific findings. We have argued that this is ill-advised for many reasons. Such elements are literally ad hoc: "for the particular end or case at hand without consideration of wider application" (*Merriam-Webster Dictionary*). This leads to the over-fitting problem discussed above, a failure to capture generalizations *within* the target domain. It also leads to failures to capture generalizations that hold *across* domains—aspects of knowledge representation, learning, and processing that are not specific to reading at all. That reading exists at all is due to the fact that it exploits capacities that evolved for other purposes. Hence our emphasis on looking for principles that govern perception, cognition, learning, and their brain bases. Advances in understanding these principles can then facilitate understanding specific tasks such as reading.

Computational modelling has contributed significantly to our understanding of word reading, but considerable work remains to be done. Even basic issues, such as how spelling-sound knowledge is represented and applied, and its relation to lexical and semantic knowledge, remain unresolved. In such a context, it is premature to focus on the degree to which existing models fit specific empirical findings. Rather, the greatest progress will come from a broader perspective that attempts to integrate the study of word reading into the more general enterprise to elucidate the neural basis of cognitive processes. In this regard, the PDP approach to modelling—quite apart from the strengths and limitations of existing models—has much to offer.

Acknowledgements

The authors are grateful to Karalyn Patterson, Marcus Taft, and Sally Andrews for comments on an earlier version.

References

Andrews, S. (1989). Frequency and neighborhood effects on lexical access: Activation or search? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 802–814.

Andrews, S. (1992). Frequency and neighborhood effects on lexical access: Lexical similarity or orthographic redundancy? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 234–254.

Andrews, S., & Scarratt, D. R. (1998). Rule and analogy mechanisms in reading nonwords: Hough dou peapel rede gnew wirds? *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 1052–1086.

Balota, D. A., Cortese, M. J., & Sergent-Marshall, S. D. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, *133*, 283–316.

Blazely, A. M., Coltheart, M., & Casey, B. J. (2005). Semantic impairment with and without surface dyslexia: Implications for models of reading. *Cognitive Neuropsychology*, *22*, 695–717.

Boyle, G. J., & Saklofske, D. H. (2004). *Psychology of individual differences* (Vols 1–4). Thousand Oaks, CA: Sage.

Browman, C. P., & Goldstein, L. (1990). Representation and reality: Physical systems and phonological structure. *Journal of Phonetics*, *18*, 411–424.

Caramazza, A. (1986). On drawing inferences about the structure of normal cognitive systems from the analysis of patterns of impaired performance: The case for single-patient studies. *Brain and Cognition*, *5*, 41–66.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Clark, H. H. (1969). Linguistic processes in deductive reasoning. *Psychological Review*, *85*, 59–108.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*, 407–428.

Coltheart, M. (1978). Lexical access in simple reading tasks. In G. Underwood (Ed.), *Strategies of information processing* (pp. 151–216). London: Academic Press.

Coltheart, M. (1999). Modularity and cognition. *Trends in Cognitive Sciences*, *3*, 115–120.

Coltheart, M. (2000). Dual routes from print to speech and dual routes from print to meaning: Some theoretical issues. In A. Kennedy, R. Radach, J. Pynte, & D. Heller (Eds.), *Reading as a perceptual process*. Oxford: Elsevier.

Coltheart, M. (2004). Brain imaging, connectionism, and cognitive neuropsychology. *Cognitive Neuropsychology*, *21*, 21–25.

Coltheart, M. (in press). Acquired dyslexias and the computational modeling of reading. *Cognitive Neuropsychology*.

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

Coltheart, M., Davelaar, E., Jonasson, J., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535–555). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

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., Sartori, G., & Job, R. (1987). *The cognitive neuropsychology of language*. Hove, UK: Lawrence Erlbaum Associates, Ltd.

Cortese, M. J., & Simpson, G. B. (2000). Regularity effects in word naming: What are they? *Memory and Cognition*, *28*, 1269–1276.

Derouesne, J., & Beauvois, M. F. (1985). The "phonemic" stage in the non-lexical reading process: Evidence from a case of phonological alexia. In K. E. Patterson, M. Coltheart, & J. C. Marshall (Eds.), *Surface dyslexia*. Hove, UK: Lawrence Erlbaum Associates Ltd.

Dunn, J. C., & Kirsner, K. (Eds.) (2003). Forum on "What can we infer from double dissociations", *Cortex*, *39*, 129–202.

Feyerabend, P. (1975). *Against method*. London: Verso.

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. Cambridge, MA: MIT Press.

Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, *5*, 674–691.

Grainger, J., & Jacobs, A. M. (1998). On localist connectionism and psychological science. In J. Grainger & A. M. Jacobs (Eds.), *Localist connectionist approaches to human cognition* (pp. 1–38). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review*, *106*, 491–528.

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.

Jared, D. (1997). Spelling-sound consistency affects the naming of high-frequency words. *Journal of Memory and Language*, *36*, 505–529.

Jared, D. (2002). Spelling-sound consistency and regularity effects in word naming. *Journal of Memory and Language*, *46*, 723–750.

Jared, D., McRae, K., & Seidenberg, M. S. (1990). The basis of consistency effects in word naming. *Journal of Memory and Language*, *29*, 687–715.

Juola, P., & Plunkett, K. (2000). Why double dissociations don't mean much. In G. Cohen, R. Johnston, & K Plunkett (Eds.), *Exploring cognition: Damaged brains and neural networks* (pp. 111–172). New York: Psychology Press.

Kuhn, T. (1962). *The structure of scientific revolutions*. Chicago: University of Chicago Press.

Lakatos, I. (Ed.) (1970). *Criticism and the growth of knowledge*. Cambridge: Cambridge University Press.

Lambon-Ralph, M. A., Patterson, K., & Graham, N. (2003). Homogeneity and heterogeneity in mild cognitive impairment and Alzheimer's disease: A cross-sectional and longitudinal study of 55 cases. *Brain*, *126*, 2350–2362.

Lichtheim, L. (1885). On aphasia. *Brain*, *7*, 433–484.

Marshall, J. C., & Newcombe, F. (1973). Patterns of paralexia: A psycholinguistic approach. *Journal of Psycholinguistic Research*, *2*, 175–199.

Massaro, D. W. (1989). Testing between the TRACE model and the fuzzy logical model of speech perception. *Cognitive Psychology*, *21*, 398–421.

McCann, R., & Besner, D. (1987). Reading pseudohomophones: Implications for models of pronunciation assembly and the locus of word-frequency effects in naming. *Journal of Experimental Psychology: Human Perception and Performance*, *13*, 14–24.

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

Newell, A. (1973). You can't play 20 questions with nature and win. In W. G. Chase (Ed.), *Visual information processing* (pp. 283–308). New York: Academic Press.

Paap, K., & Noel, R. W. (1991). Dual route models of print to sound: Still a good horse race. *Psychological Research*, *53*, 13–24.

Patterson, K. E., Marshall, J. C., & Coltheart, M. (1985). *Surface dyslexia: Neuropsychological and cognitive studies of phonological reading*. Hove, UK: Lawrence Erlbaum Associates Ltd.

Patterson, K., Lambon-Ralph, M. A., Jefferies, E., Woollams, A., Jones, R., Hodges,

J. R., et al. (2006). "Pre-semantic" cognition in semantic dementia: Six deficits in search of an explanation. *Journal of Cognitive Neuroscience*, *18*, 169–183.

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

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

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

Plaut, D. C., & McClelland, J. L. (2000). Stipulating versus discovering representations: Commentary on M. Page, Connectionist modeling in psychology: A localist manifesto. *Behavioral and Brain Sciences*, *23*, 489–491.

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

Putnam, H. (1973). Reductionism and the nature of psychology. *Cognition*, *2*, 131–146.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*, 59–108.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*, 372–422.

Reichle, E. D., Pollatsek, A., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, *105*, 125–157.

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

Sears, C. R., Hino, Y., & Lupker, S. J. (1995). Neighborhood size and neighborhood frequency effects in word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 876–900.

Seidenberg, M. S. (1985). The time course of phonological code activation in two writing systems. *Cognition*, *19*, 1–10.

Seidenberg, M. S. (1988). Cognitive neuropsychology and language: The state of the art. *Cognitive Neuropsychology*, *5*, 403–426.

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

Seidenberg, M. S., & Plaut, D. C. (1998). Evaluating word-reading models at the item level: Matching the grain of theory and data. *Psychological Science*, *9*, 234–237.

Seidenberg, M. S., Plaut, D. C., Petersen, A. S., McClelland, J. L., & McRae, K. (1994). Nonword pronunciation and models of word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 1177–1196.

Seidenberg, M. S., Waters, G. S., Barnes, M. A., & Tanenhaus, M. K. (1984). When does irregular spelling or pronunciation influence word recognition? *Journal of Verbal Learning and Verbal Behavior*, *23*, 383–404.

Shallice, T. (1988). *From neuropsychology to mental structure*. Cambridge: Cambridge University Press.

Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review*, *81*, 214–241.

Spieler, D. H., & Balota, D. A. (1997). Bringing computational models of word naming down to the item level. *Psychological Science*, *8*, 411–416.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, *30*, 276–315.

Taft, M., & Russell, B. (1992). Pseudohomophone naming and the word frequency effect. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *45A*, 51–71.

Taraban, R., & McClelland, J. L. (1987). Conspiracy effects in word recognition. *Journal of Memory and Language*, *26*, 608–631.

Thomas, M. S. C., & Karmiloff-Smith, A. (2003). Connectionist models of development, developmental disorders and individual differences. In R. J. Sternberg, J. Lautrey, & T. Lubart (Eds.), *Models of intelligence: International perspectives* (pp. 133–150). Washington, DC: American Psychological Association.

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

Waters, G. S., & Seidenberg, M. S. (1985). Spelling-sound effects in reading: Time course and decision criteria. *Memory and Cognition*, *13*, 557–572.

Woollams, A., Lambon-Ralph, M. A., Hodges, J. R., & Patterson, K. (2005). *SD-squared: On the association between semantic dementia and surface dyslexia*. Paper presented to the Experimental Psychology Society meeting, April 2005.

Zevin, J. D., & Seidenberg, M. S. (2006). Simulating consistency effects and individual differences in nonword naming: A comparison of current models. *Journal of Memory and Language*, *54*, 145–160.

Notes

1 Note that we are not asserting that a model has to be tested against every study of a given phenomenon that has appeared in the literature. Experiments often yield varying results, for a variety of reasons. Some studies are clear outliers, insofar as they fail to replicate, or are contradicted by, multiple other studies. Some studies yield results that may be valid but have not been replicated, and so their status is unclear. It is appropriate to omit outliers and unclear cases in testing a model. However, it would not be appropriate to select studies on the basis of whether a model simulates them correctly or not. The safest strategy is to focus on robust phenomena that have replicated in multiple studies.

2 Coltheart et al.'s simulation of Jared's (1997) study is problematic. They asserted that consistency effects in studies such as hers were due to two confounding factors: the inconsistent words included some items that DRC treats as exceptions, and included more words that create "whammies", temporary misapplications of rules based on a left-to-right pass through the word. However, when the suspect items are removed, the consistency effect remains in Jared's data, but not in the DRC simulation.

3 The concept of "localist representation" engenders some confusion. A representation is localist or distributed only with respect to a specific set of entities. For example, in McClelland and Rumelhart's interactive activation model, the representations at the letter level are localist with respect to letters but distributed with respect to words; the same is true of DRC. A PDP model such as that of Harm and Seidenberg (2004) used localist representations of letters in orthography, but distributed representations for words in orthography, phonology, and semantics. The contrast is not between models employing localist versus distributed representations, since all of the above models include both. Rather,

DRC is committed to the narrower claim that there are localist representations of *words*.

4 The usefulness of the distinction between descriptive and explanatory adequacy is distinct from questions concerning the extent to which Chomsky's own theories achieve these goals.