COGNITIVE NEUROPSYCHOLOGY, 2010, 27 (6), 477–494

Psychology Press
Taylor & Francis Group

Case series investigations in cognitive neuropsychology

Myrna F. Schwartz[1] and Gary S. Dell[2]

[1]Moss Rehabilitation Research Institute, Elkins Park, PA, USA
[2]Beckman Institute, University of Illinois at Urbana-Champaign, Urbana, IL, USA

Case series methodology involves the systematic assessment of a sample of related patients, with the goal of understanding how and why they differ from one another. This method has become increasingly important in cognitive neuropsychology, which has long been identified with single-subject research. We review case series studies dealing with impaired semantic memory, reading, and language production and draw attention to the affinity of this methodology for testing theories that are expressed as computational models and for addressing questions about neuroanatomy. It is concluded that case series methods usefully complement single-subject techniques.

Keywords: Aphasia; Case series; Lexical access; Semantic memory; Computational models.

To many in the broader cognitive neuroscience community, "cognitive neuropsychology" (cn) is identified with a rigorous single-subject methodology. Leaders in the field have long promoted this view. For example, in their introduction to the 20th anniversary issue of *Cognitive Neuropsychology*, Alfonso Caramazza and Max Coltheart state, "It is deeply characteristic of cognitive neuropsychology that it studies symptoms rather than syndromes and carries out single case studies rather than group studies" (Caramazza & Coltheart, 2006, p. 5). This pronouncement is supported by the special issue articles, which summarize the impressive contributions of cn research to many different domains of cognition; in virtually every domain, single-subject research features prominently. Yet, to many who practise cognitive neuropsychology, the field's identification with single-subject research is overstated, and the rhetoric on the topic is unnecessarily contentious. In journals devoted to cognitive psychology, neuropsychology, and cognitive neuroscience, one can easily find high-profile cn papers that report data from sizeable collections of patients. Just like single-subject studies, those testing several patients are carried out to draw inferences about the functional organization of cognition. The most successful of these, in our view, have used the case series approach. Our goal in this article is to show why, as an alternative and complement to the study of individual cases,

Correspondence should be addressed to Myrna F. Schwartz Moss Rehabilitation Research Institute, 50 Township Line Rd., Elkins Park, PA 191027, USA. (Email:mschwart@einstein.edu).

We acknowledge with gratitude the support of NIH/NIDCD (National Institutes of Health/National Institute on Deafness and Other Communication Disorders) Grant RO1DC000191–28 to M.F.S. and helpful suggestions from Dan Mirman and Simon Fischer-Baum on an earlier version of the manuscript.

http://www.psypress.com/cogneuropsychology DOI:10.1080/02643294.2011.574111

case series methodology is important to cn and will probably contribute even more to the cn of the future.

INTRODUCTION TO THE CASE SERIES DESIGN

In clinical medicine and epidemiology, the case series design is a recognized alternative to cohort and case-control designs. Participants are followed for a period of time with the same, relevant data recorded from each. There is no control group, and data from the sample are not aggregated. Instead, the target event (e.g., disease onset) is modelled in relation to patient, time, and/or treatment variables using regression techniques (e.g., Farrington, Nash, & Miller, 1996).

The situation is similar in cn case series investigations. For the target cognitive ability, one studies a sample of brain-injured patients who are expected to vary on that ability, obtaining uniform performance measures on the target and potentially related clinical, anatomical, and performance measures and analysing how the measures covary. The goal of the analysis is to understand the cognitive mechanisms responsible for the covariance, and this often involves developing or testing a specific statistical or processing model. The patient sample may be defined broadly (e.g., chronic aphasia; Schwartz, Dell, Martin, Gahl, & Sobel, 2006) or narrowly (e.g., dysgraphic individuals with lexical and sublexical spelling deficits; Fischer-Baum & Rapp, 2008). As well, the assessment given to each patient may vary in its breadth and depth. Such methodological decisions are determined by the study question, the maturity of relevant theory, and the usual practical trade-offs.

Case series investigations in cn should be distinguished from the many neuropsychological studies in which patient group means are of primary interest. These kinds of studies treat within-group variability on some dependent variable as a source of noise, effectively removing that variability as an object of consideration. A case series, in contrast, preserves and uses the individual data by characterizing the distribution of scores and, most importantly, what factors covary with the scores. Of course, because there are several patients in a case series, one can group them in various ways for purposes of description and even for statistical inference. The grouping can be done on an a priori basis (Jefferies & Lambon Ralph, 2006). Alternatively, the groups could be assembled based on scores from the case series data itself, provided that inferential tests are pursued with great caution given the potential for contamination from "peeking" at the data. In either event, grouping can help the researcher see the patterns in a case series.

Case series methodology also should be distinguished from single- or multiple-case methods. Multiple-case studies are concatenations of two or more related case studies. Key differences can be illustrated by considering an early case series report authored by Patterson and Hodges (1992). The report featured six individuals with a history of semantic memory loss, as demonstrated on tests of naming and comprehension. Noted at the outset was the report's "unusual format: it is neither a group study, nor a typical single-case study, nor even a series of complete single-case studies. Rather it represents a focus on one specific aspect of six single-case studies" (p. 1026). That focus was on how the reading of regular and exception words was impacted by two independent variables: severity of the semantic deficit and word frequency. Their analysis showed that lexical frequency had a dramatic impact on the subjects' reading of exception words but not regular words, and that the exception-regular difference was greater for those with severe semantic loss than for the milder subjects.

This "proto" case series lacked some elements that are now considered important: a sample size suitable for identifying and parameterizing linear and more complex trends in the data (usually an n of 10 or more) and uniform data gathering on each patient. Nevertheless, it well exemplifies that the essence of a cn case series is the analysis of patient variation in relation to other, theoretically relevant variables. Consistent with this goal,

and in contradistinction to the multiple-case study, the cognitive analysis tends to be circumscribed. Typically, tests are administered or reported only if they bear on the selection criteria or the study hypotheses—that is, there is not a systematic attempt to fully characterize the deficits of each individual. There is no principled reason why case series analyses cannot be combined with intensive cognitive characterization of each patient. In practice, though, such characterization tends to be reserved for patients who deviate from the group trends, in order to track down the reason for the deviation (e.g., Schwartz et al., 2006).

In a notable departure from cn orthodoxy, case series studies often define the sample clinically (e.g., poststroke aphasia; anomia; semantic dementia) and not purely along cognitive lines. The traditional objection to clinical classifications is, of course, that they invite unhelpful heterogeneity with respect to the cognitive ability in question. Because case series have as their goal to identify and explain variation, some degree of heterogeneity is deemed welcome, and even necessary. As a general rule, one can say that proponents of cn case series are willing to relax the restrictions on heterogeneity in the exploration of theoretically interesting patient variation. We have more to say about this in subsequent sections.

Prior discussions of the use of case series methods in cn have tended to draw the contrast with the single-subject approach differently from the way we do here. For example, Patterson and Plaut (2009) associate single-subject methods with the search for functional dissociations that support modular models of cognition. They associate case series methods with the search for functional *associations* that support parallel, distributed processing models. This division reflects the historical record but does not do justice to the flexibility of either approach. For example, as discussed below, Dell and colleagues used case series methods to uncover evidence of both associations and dissociations in aphasic naming performance, evidence that was used to support a processing model that combines elements of modularity and interactivity. Similarly with the single-subject approach, the evidence used to support processing accounts need not be limited to within-subject dissociations. For example, Warrington (1975) first demonstrated the now-familiar profile of semantic memory dissolution in three intensively studied patients with progressive cortical degeneration. Each patient was shown to exhibit hierarchically structured comprehension loss for both words and pictures (an association) in the context of relatively preserved episodic memory and nonsemantic language functions (a dissociation).

Why case series?

Our central claim is that case series have an important function in cn—that there is a good reason to test a set of individuals on a common set of measures and analyse the data as a group. So, let us start with the standard objection to group studies. When comparing two or more patient groups, it is perilous to assume that members of a group have similar deficits—that is, that the group is homogeneous. The mean performance of the group may characterize few or none of its members (e.g., McCloskey & Caramazza, 1988). The traditional solution to this concern is to adopt the single-subject approach. Given this, McCloskey (1993, p. 725) then asks a cogent question:

> One could conduct patient-group studies in which each patient was tested extensively on a substantial number of tasks and results were then analyzed at both individual and group levels. However, the following question would then arise: Given the acknowledged need to consider individual patients' performance patterns, what function would be served by aggregating the data over subjects . . . ?

As indicated above, for case series studies, the function of analysing the patients together is to identify theoretically important quantitative trends in the sample, particularly those in which some model makes precise predictions about the nature of those trends. Such trends simply cannot be determined unless patients are analysed together.

Although the kind of grouping that is carried out in case series studies can lead to valuable

findings, we do not say that these studies are immune from the problems of group studies. On the contrary, any patient sample will lack homogeneity, and that will compromise interpretation of any cross-sample statistic, not just claims about sample means. For example, any group quantitative trend, such as a regression slope relating one variable to another, may vary among patients and, in so doing, may reflect multiple factors. This is especially likely when the case series sample is broadly defined, and the cognitive assessment is not extensive. We consider methods of dealing with this concern below, in "Heterogeneity in Case Series".

It is our firm belief that the case series approach is neither more nor less suited to the goals of cn than other methods. One should evaluate each case series for whether it succeeds in advancing cognitive theory based on the usual criteria: Are the measures appropriate to the question being asked and the patients being studied? Have they generated data that are valid and reliable? Have important confounds been addressed? Do the results advance knowledge about the nature of human cognition? In the following sections, we illustrate case series research in cn, by describing particular studies for which we have a good understanding of their motivations and methods, either because we were involved in the study or because we are familiar with the research domain. These can serve as examples of the kinds of research questions that are particularly suited to a case series approach and highlight methodological issues that may arise when this approach is adopted.

IDENTIFYING TRENDS AND TESTING MODELS THROUGH CASE SERIES ANALYSIS

Severity-related interactions in lexical access deficits in aphasia

Speech-error evidence from normal and aphasic speakers demonstrates that some errors index difficulty in mapping semantic representations onto lexical representations, whereas others reveal problems in retrieving and sequencing phonological segments. One can take a semantic error as indicative of the former mechanism and a nonword error as indicative of the latter. The literature reports clear examples of patients in whom nearly every error was semantic and patients in which most errors were nonwords, suggesting a qualitative distinction between the cognitive processes responsible for the two error types (see Ruml & Caramazza, 2000, for review). In a case series, though, Dell, Schwartz, Martin, Saffran, and Gagnon (1997) presented evidence that differences in semantic and nonword-error rates were, to a large extent, related to overall severity of the naming impairment. In fact, the associations between naming correctness and semantic-error proportion and between correctness and nonword-error proportion revealed that the mild patients' errors were generally semantic, while the severe patients produced many nonword responses. Figure 1 shows the severity functions for the two error types as second-degree polynomials based on the data from the 21 patients in this case series. Note that "severity" in the figure runs from high (few correct) to low (many correct).

Semantic errors occur in the relatively error-free normal pattern and increase with severity up to a point and then decrease. Nonword errors are rare in mild cases, but increase dramatically in severe cases, reaching their maximum in the most severe cases. This analysis of patient performance across the severity range suggests an alternative to the hypothesis that semantic and nonword errors arise from qualitatively distinct mechanisms. Dell et al. (1997) proposed that what underlies variation in both kinds of error is quantitative variation in global processing mechanisms that interact with the structure of the lexical access system to promote semantic errors when it is mildly damaged and nonword errors when damage is more severe. One hypothesized mechanism was the rate of activation decay; another was the rate at which activation spreads. Subsequently, these researchers modified the idea of global processing deficits because of findings

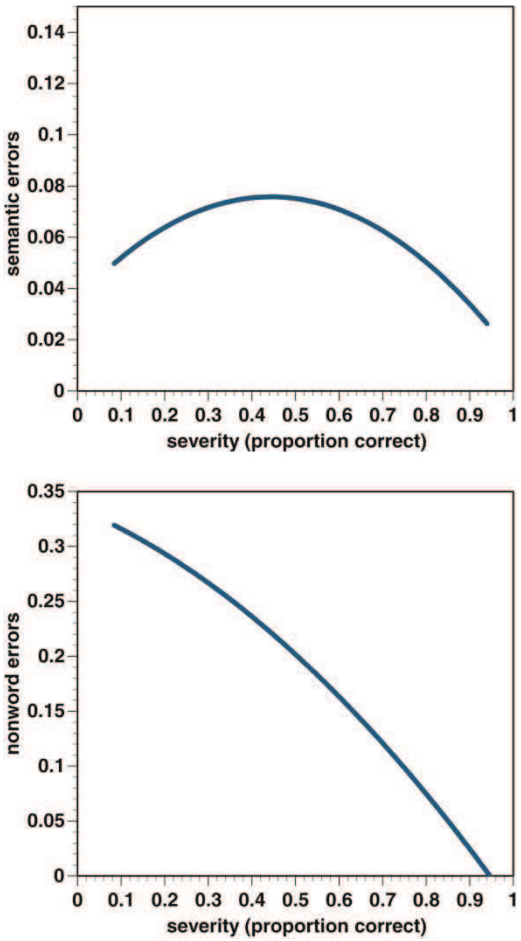

Figure 1. *Severity curves based on the 21 modelled patients in Dell, Schwartz, Martin, Saffran, and Gagnon (1997). Y-axes show error rates as proportions of all responses. Both curves are second-degree polynomials.*

from a larger case series that had the power to reveal systematic variations *within* the severity-based patterns exemplified in Figure 1 (Schwartz et al., 2006). So, on top of the severity-based patterns discovered in the earlier case series study, the larger study found, for example, patients with poor naming, but many fewer nonword errors than would be expected just from the severity functions. Later, we expand on the specific contribution of the Schwartz et al. case series. For now, we simply draw attention to the goals and the methods of the case series in Dell et al. (1997): It aimed to understand lexical access by examining how the pattern of different naming errors (semantic and nonwords) varies among aphasic subjects, and particularly how the error pattern changes with severity. It did so by identifying quantitative trends in the data using regression, specifically a nonlinear regression model as it became clear that the relations between some error types and severity were often complex, such as that shown with the semantic errors in Figure 1.

Statistical versus processing models in case series

We have said that the goal of a case series is to explain the variation in the primary measures taken from a patient sample in order to draw inferences about cognitive functions. This explanation typically takes the form of a quantitative model. Often, as illustrated in the previous section, that model is a statistical model, such as regression. One or more dependent variables in the case series are predicted from other measures or patient characteristics by constructing prediction equations. The form and coefficients of these equations are then interpreted in the light of theory. For example, this has been the preferred strategy for determining how the naming performance (accuracy and/or error scores) of a neuropsychological population is influenced by target properties such as frequency, length, and age of acquisition (e.g., Kittredge, Dell, Verkuilen, & Schwartz, 2008; Lambon Ralph, Graham, Ellis, & Hodges, 1998; Nickels, 1995; Nickels & Howard, 1994, 1995, 2004; Woollams, Cooper-Pye, Hodges, & Patterson, 2008). Care should be taken, though, in using statistical techniques that only consider linear relations among the measures (e.g., linear regression, principal components analysis). The example in Figure 1 comparing semantic and nonword naming errors illustrates the need for a consideration of nonlinear and especially non-monotonic relationships between measures. The ability to detect such relationships is a particular strength of large case series studies (e.g., Woollams et al., 2008). But curvilinear

associations can easily be missed if researchers do not consider the possibility.[1]

Ultimately, the best tool for understanding the complex relations among the measures taken in a case series may be an explicit processing model rather than a statistical model such as regression. This is especially important for case series in cn, which uses the analysis of deficits to draw inferences about unimpaired processing (see Dell & Caramazza, 2008; Goldrick, 2008, for discussion of models in cn).

A processing model can be computationally implemented to mimic the operations that underlie the tasks being performed. The best models will directly generate analogues to the real data. So, a model of deep dyslexia will sometimes make, say, a semantic error when attempting to "read" a concrete word (e.g., Plaut & Shallice, 1993). Moreover, different versions of the model can be set up to represent relevant between-patient differences—for example, in lesion severity, distribution, or premorbid cognitive capacity. Such a model is then poised to simulate the functional relations between the patient measures. We exemplify the coupling of case series methods and computational processing models with published work in the areas of semantic memory, reading, and lexical access.

Semantic memory

The influential "distributed-plus-hub" theory of semantic memory (Lambon Ralph, Sage, Jones, & Mayberry, 2010; Patterson, Nestor, & Rogers, 2007; Rogers, Lambon Ralph, Garrard et al., 2004) derives support from three empirical pillars:

1. Case series data from patients diagnosed with semantic dementia (SD; e.g., Bozeat, Lambon Ralph, Patterson, Garrard, & Hodges, 2000) or with the fluent form of primary progressive aphasia (Adlam et al., 2006) reveal common patterns of covariation within and across verbal and nonverbal semantic tasks. The hub theory takes this commonality as evidence for amodal semantic representations and their vulnerability to disruption.
2. Studies that correlate the multimodal semantic symptoms of SD with in vivo measurement of brain atrophy and blood flow have highlighted the importance of the inferior and lateral aspects of the anterior temporal lobes (Mummery et al., 2000; Mummery, Patterson, Wise, Vandenbergh, Price, & Hodges, 1999; Rogers et al., 2006) and have led to the view that this is the "hub" that houses amodal semantic representations.
3. A parallel, distributed processing (PDP) implementation of the hub theory has been developed (Rogers, Lambon Ralph, Garrard et al., 2004), within which the hub functions to abstract the similarity structure across the multiple, modality-specific knowledge representations with which it interacts. We now describe the Rogers, Lambon Ralph, Garrard, et al. (2004) semantic memory model and show how its development and evaluation are inextricably linked with case series data.

The model's architecture features one pool of localist units representing verbal descriptors (names and perceptual, functional, and encyclopaedic propositions), another pool representing visual features, and recurrent semantic ("hidden") units that interconnected with the verbal and visual units via bidirectional connections whose weights were set by learning. The recurrent semantic units constitute the amodal hub.

[1]A related concern about nonlinearity arises because most neuropsychological tests yield a percentage value—that is, a value on a scale from zero to a logical maximum such as 100%. Analyses that treat percentage score differences as equivalent across the scale may be misleading. Logistic regression, in which percentages are transformed into the logit of the response types—that is, ln[p(correct)/p(incorrect)]—is considered an appropriate remedy for this problem in percentage/proportion data (Jaeger, 2008), and we recommend that these methods be adopted where appropriate in neuropsychological case series analysis (Dilkina, McClelland, & Plaut, 2008). Moreover, with the advent of new software, logistic regression can now be done in a multilevelled manner, so that each measurement trial can be associated with a specific participant and item, and both participants and items can be treated as random effects (e.g., Nozari, Kittredge, Dell, & Schwartz, 2010).

A training environment was constructed to capture the similarity structure of items identified in published attribute-norming experiments (Garrard, Lambon Ralph, Hodges, & Patterson, 2001; McRae, de Sa, & Seidenberg, 1997). The model was trained on visual and verbal patterns until it learned the associations among the names, appearances, and descriptions of items in its virtual environment. Items were assigned unique names that were either general (e.g., "bird", "tool") or specific ("robin", "drill"); unique descriptors were also either general (e.g., living, nonliving) or more specific (mammal, tool, fruit).

In the test phase, simulations were run on analogues of four semantic memory tests used with actual SD patients. Picture naming was tested in the model by inputting a prototypical visual pattern (representing a pictured target) and requiring as output the unique name (general or specific). Word and picture sorting were tested by inputting a name or a visual pattern and requiring as output the correct verbal descriptor (general or specific). For word–picture matching, the model was presented with a name followed by a series of visual inputs corresponding to the target and distractor pictures. The model's "choice" was defined as the visual input that generated within the semantic units an internal state most similar to that produced by the target name. Finally, to simulate drawing to command, the model was given a name as input, and its response was defined by the resulting pattern of activity across the visual units.

On each task, Rogers, Lambon Ralph, Garrard, et al. (2004) compared the model's performance under varying degrees of impairment severity (simulated by removing an increasing proportion of the semantic weights) with the data from actual patients, which were similarly described in relation to severity. In this way, the manipulation of severity provides the link between the modelling and the case series data. It is the hypothesized source of variance in the simulated and the actual behaviour. For example, as naming accuracy decreased (i.e., severity increased), the proportion of semantically related substitutions declined, while the proportion of superordinate responses and omission increased. In addition, sorting performance declined with increasing deficit of semantic impairment (as measured by word–picture matching), with specific-level sorts (e.g., animals and tools) suffering a steeper decline than more general-level sorts (living and nonliving). These severity-related interactions were shown to be comparable in the model and the patients. Model and patients also performed similarly in other theoretically relevant respects. In naming, for example, the likelihood of omissions was greater for artifacts than animals, whereas semantic and superordinate errors were more likely with animals. In drawing, model and patients were more likely to omit distinctive features of objects, whereas they were more likely to inappropriately add features that were widely shared.

The model's simulations prompted Rogers, Lambon Ralph, Garrard, et al. (2004) to conclude that the functional explanation for these and other features of semantic dementia lies in the structure that the semantic system forms through learning, coupled with the dynamics of processing as knowledge degrades. Reflecting the structure in the environment, the model develops a dense semantic space (or "neighbourhood") for animals, whereas the space for artefacts is sparse. A dense semantic neighbourhood affords more opportunities for the system to be captured by the wrong attractor, which explains why, in naming, semantic errors are more likely for animals whereas omission errors are more likely for artefacts. Generally speaking, as the semantic system degrades, it has a harder time distinguishing among concepts. In a mildly damaged system, the network may settle into an inappropriate attractor, from which it might be unable to produce distinguishing information (that zebras have stripes) but might succeed in producing information of a more general nature (yielding errors such as zebra → horse or zebra → animal). With greater damage, the only information available might be information that is common to many items in the domain (superordinate response) or to no known entity (omission). Later empirical work has supported this functional account by showing that the typicality of an item within its

semantic category strongly impacts SD patients' performance on a variety of tasks, not all of which have an obvious semantic component (Patterson et al., 2006; Rogers, Lambon Ralph, Hodges, & Patterson, 2004; Woollams et al., 2008).

Surface dyslexia

The best known of the "nonsemantic" typicality effects in semantic dementia is surface dyslexia, which features the tendency to regularize exception words in reading aloud (e.g., "pint" pronounced as /pɪnt/). Earlier we mentioned Patterson and Hodges's (1992) proto case series, which first posited a causal relation between loss of meaning and impaired exception word reading. The argument has been bolstered with case series data from larger SD samples (Graham, Patterson, & Hodges, 1995; Patterson et al., 2006; Woollams, Lambon Ralph, Plaut, & Patterson, 2007). In the Woollams et al. (2007) study, data from 51 patients (100 data points, since some were tested at more than one point in time) confirmed earlier reports of a quantitative relation between the semantic and reading deficits with the impressive finding that a composite semantic measure explained half of the variance in exception word reading. At the same time, the study confirmed prior evidence that there are occasional patients who violate this association (Blazely, Coltheart, & Casey, 2005; Cipolotti & Warrington, 1995; Schwartz, Saffran, & Marin, 1980), here exemplified by 3 patients who showed unexpectedly good reading, given their semantic scores. Pointing to the fact that all 3 did develop surface dyslexia as their SD progressed, the authors rejected the view that the occasional dissociation of reading performance and semantic status indicates that semantics and word pronunciation are functionally independent (Blazely et al., 2005; Coltheart, 2006; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001). Instead, they argued that the overall association between semantics and reading and the occasional dissociations between them could be explained in a single functional account that incorporates semantic mediation of word pronunciation and individual differences in the degree of reliance on this mediating input.

The proposed functional account is based on the PDP triangle model (orthography, phonology, semantics) and its phonological-semantic "division of labour". This refers to the fact that as the model is trained to pronounce words, the phonological pathway (orthography to phonology) becomes specialized for reading words with frequent and/or consistent mappings (i.e., regular words), whereas the semantic pathway (semantics to phonology) assumes importance for the more arbitrary mappings that characterize exception words. Simulating surface dyslexia by diminishing the strength of the semantic contribution negatively impacts the model's performance with exception words, particularly those of low frequency (Harm & Seidenberg, 2004; Plaut, McClelland, Seidenberg, & Patterson, 1996).

To explain the apparent independence of semantics and reading in some patients, Plaut (1997) introduced a variant of the triangle model that simulated premorbid individual differences in how much semantic activation is supplied during the training phase. Woollams et al. (2007) implemented a similar model to achieve multiple instantiations of an intact network with individual differences in semantic reliance. The multiple instantiations were then aggregated to investigate the impact of lesioning the semantic input to phonology, which they achieved by progressively decreasing activation and adding noise. The model's performance was compared to the empirical, case series data, which were analysed cross-sectionally and longitudinally. In both the models and patients, slopes of the function relating lesion severity/semantic loss to reading accuracy exceeded zero for all types of words, but not for nonwords. Consistent with surface dyslexia, slopes were steeper for low- than for high-frequency exception words and steeper for exception words than for regular words. Outliers, relative to best fit lines, were rare. Longitudinal analysis of the patient data revealed similar effects of worsening semantic status. Furthermore, as we noted earlier, the 3 patients

who, on first testing, qualified as having unimpaired reading in the presence of semantic loss all developed surface dyslexia as their semantic dementia worsened. Overall, there was an impressive match between the performance of the patients and the individual differences triangle model.

In a subsequent development, Dilkina et al. (2008) merged features of the individual differences triangle model (Plaut, 1997; Woollams et al., 2007) with the Rogers, Lambon Ralph, Garrard, et al. (2004) semantic memory model described earlier, in order to simulate both the reading and picture naming performance of 5 patients (Figure 2). Individual differences were incorporated into the model, representing premorbid differences in the effectiveness of the direct (orthography to phonology) reading pathway and postmorbid differences in spatial distribution of the brain atrophy. The resulting model gave a good account of the positive, curvilinear relation between reading and picture naming that was present in the data. Critically, it also explained the behaviour of those patients whose naming and reading were dissociated, such that reading was unexpectedly spared in relation to naming (e.g., patient E.M. in Blazely et al., 2005). It did this by showing comparable performance in versions of the model that departed from the norm on one or more individual difference parameters.

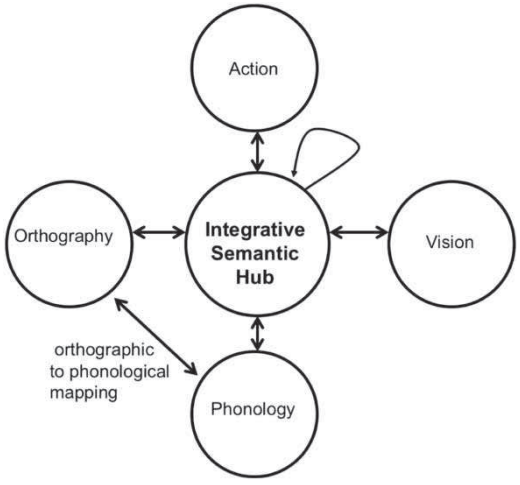

Figure 2. *Based on Dilkina, McClelland, and Plaut (2008). The model combines features of Rogers et al.'s (2004) semantic memory model with the triangle model of reading. The integrative semantic hub consists of hidden units that connect to one another and also bidirectionally with each of the four input/output levels (action, orthography, etc.). The direct orthographic to phonological mapping becomes specialized for the pronunciation of regular words, while the semantic hub plays a stronger role for exception words. Semantic damage leads to both semantic dementia and surface dyslexia.*

While neither this nor the Woollams et al. (2007) study rules out the possibility that semantics and reading are functionally dissociable (note that neither study provides empirical evidence that the patients actually differed in reading experience, etc.), the authors of these studies give a good account of their case series data by putting individual differences into the semantic-mediation theory of reading. Spurred by these seminal studies, future case series investigations are likely to include much more in the way of individual difference data and to feature these in the processing accounts they offer.

The studies of semantic memory and surface dyslexia illustrate the value of linking up a case series with a processing model. The model can offer predictions about quantitative trends regarding how performance can vary (and even the shape of the function, e.g., Dilkina et al., 2008) and possible mechanisms for this variation (e.g., individual differences in reading experience). It is the focus of the case series on quantitative accounts of the variation that makes the data of interest and particularly suited to model development and testing.

Lexical access in naming

Our final example of the valuable link between case series and processing models returns to the study by Dell et al. (1997) on aphasic picture naming. This study was carried out to test the two-step interactive model of production (Dell & O'Seaghdha, 1992). Like the two previous models presented here, it performs lexical tasks by spreading activation in a network and can simulate individual trials resulting in either correct responses or various kinds of errors (e.g., semantic or nonword errors). However, the interactive two-step model associates brain damage with changes in model processing parameters, rather than the

removal of network units and connections. For example, damage in such models can be attributed to increases in activation noise (e.g., Rapp & Goldrick, 2000) or activation decay rate (Dell et al., 1997; Martin, Dell, Saffran, & Schwartz, 1994). The most recent version of the model attributes aphasic variation to two parameters: s-weight, which represents the strength of the connections between semantic and lexical units; and p-weight, the strength between phonological and lexical units (Schwartz et al., 2006). Patients are assigned parameter values by a fitting process (Dell, Lawler, Harris, & Gordon, 2004). First, the patient's response profile in the naming task is determined—for example, .78 correct, .03 semantic errors, .03 phonologically related word (formal) errors, .02 mixed semantic–phonological errors, .01 unrelated word errors, and .13 nonword errors. Then the model attempts to generate this pattern by varying its s- and p-weights. In this case, the model would find an excellent fit with an s-weight of .027 and a p-weight of .019, indicating that this case is somewhat more impaired in lexical–phonological mappings than in semantic–lexical mappings. Notice that, in the context of a case series study, the model's parameters can themselves then act as dependent variables in the series, and their distribution and association with other measures can be assessed. For example, Dell, Martin, and Schwartz (2007) demonstrated that the p-weight determined from picture naming strongly predicts word repetition performance, thus suggesting that the lexical–phonological component of naming is shared with repetition (see also Nozari et al., 2010).

Variations in model parameters can explain how measures vary with severity in the case series. Recall that semantic and nonword naming errors behave quite differently across severity (see Figure 1). Varying model parameters generates a severity continuum, and thus it can be seen whether the model simulates this behaviour. Figure 3 shows an example in which weights vary from their normal to nonfunctional values. Just as in the data (see Figure 1), the model's semantic errors show a nonmonotonic relation with severity, while nonword errors are clearly

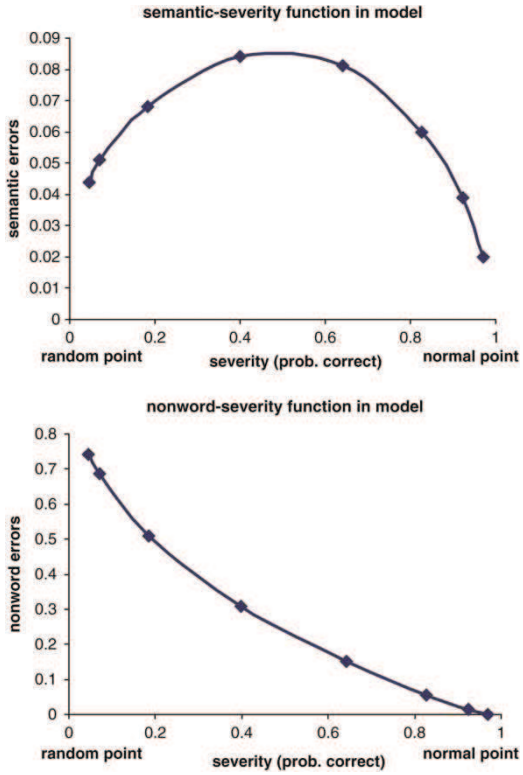

Figure 3. *For the interactive two-step model, semantic and nonword error probability (relative to total responses) as a function of the proportion of correct responses (severity). Compare to Figure 1. The points of differing severity were created by reducing both s- and p-weights by the same proportions.*

associated with the more severe cases. Why does the model work this way? It has to do with differences between the model's normal behaviour and its behaviour when it is completely broken down and thus generating only "random" responses. The model's normal error pattern was deliberately set up to match the error profile of normal controls. In this profile (called the "normal point"), errors are rare, but semantic errors are the dominant error. At the other end of the severity continuum, the model was set up to match the distribution of errors that would result from the production of random word-length strings that are nonetheless phonotactically legal. This is called the "random point". In English, most such strings would be nonwords, but some would be words (most of which would be unrelated to the

target). Thus, the model defines a space of possible error patterns, a space that runs from the normal to the random point. The reason that semantic errors increase and then decrease as damage becomes more severe is largely a consequence of the rarity of semantic errors at the random point. Nonword errors behave differently because they are nonexistent at the normal point, but overwhelmingly predominant at the random point. In essence, the model realizes the continuity thesis of aphasia (Dell et al., 1997; Freud, 1891/1953) that quantitative variation between normality and randomness defines the possible deficits. However, it is important to recognize that the model does not attribute everything to severity. According to the model, the aphasic space has two dimensions, one for s-weight and one for p-weight. Thus, there is a qualitative aspect (kind of damage) along with quantitative degree of damage that emerges from the application of the model to the case series data. The combination of model and case series was thus crucial for sorting out the quantitative and qualitative sources of variation in naming performance by aphasic speakers.

We noted that the model defines a space of possible error patterns. If the model is a good one, it should be able to fit the naming data from any patient in the target population—namely, post-stroke aphasia. In Schwartz et al. (2006), naming data were obtained from 94 individuals with diverse presentations of chronic aphasia. In the large majority of cases, the model provided a good fit to the observed error pattern. Schwartz et al. then took a closer look at patients whose error pattern deviated from that of the model by a criterion amount, in effect treating these patients as single cases. Most of these patients ($N = 9$) exhibited what was called the "pure semantic" pattern, meaning that their errors were overwhelmingly semantic, despite a below-normal level of accuracy. This error pattern is unexpected by the model, because a breakdown in the s-weight that produces many semantic errors tends also to create formal and unrelated errors. Schwartz et al. were able to explain some, but not all, of these deviating patients. For example, further testing showed that some deviators had an additional central semantic deficit. Consequently, some of their semantic errors may have arisen from poor conceptualization. The model, which assumes correct semantic input as a simplifying assumption, thus does not take note of these errors. The deviations from other patients could be attributed to another simplification, the treatment of omission errors (e.g., Dell et al., 2004). Ultimately, though, there were 3 patients with the pure semantic pattern that could not be explained by these factors. These patients thus present a challenge to the model's underlying theory, possibly to the model's key assumption that activation flows from lower levels (e.g., phonemes) to higher levels (words and semantics) with as much strength as it does in the top-down direction (Goldrick & Rapp, 2002; Rapp & Goldrick, 2000; Ruml, Caramazza, Capasso, & Miceli, 2005).

We have said that the strength of a model/case-series combination is the potential to test precise predictions about patient variation. For example, in the interactive two-step model, formal errors in naming (e.g., "mat" or "sat", for "cat"), have a double nature. Formal errors can be caused either by the selection of the wrong word during the step associated with lexical access, or by the incorrect phonological encoding of the correct word during a later step when that selected word's phonemes are chosen. If these errors occur during lexical access, they must, according to the model, match the grammatical class of the target (in object picture naming, they must be nouns, e.g., "mat" for "cat"), but if they are phonological, they could make non-nouns (e.g., "sat"). It turns out that, in the model, the extent to which such errors are lexical as opposed to phonological is directly proportional to the value (p − s)—that is, the extent to which the model's assigned phonological parameter is stronger than the semantic parameter.[2]

[2]This is because formal errors at the lexical level require a sufficiently low semantic parameter so that lexical selection of the target is impaired, but a sufficiently strong phonological parameter that phonological feedback activates formal competitors (such as "mat"), and that, if "mat" is selected, it is correctly pronounced.

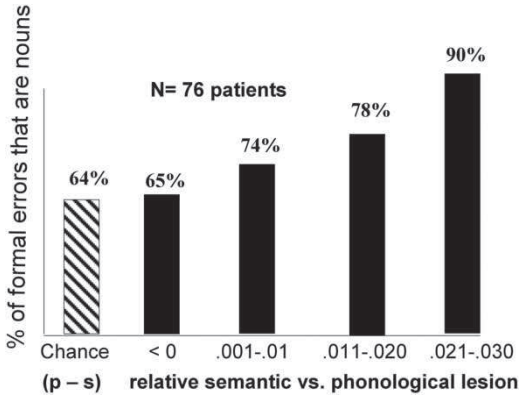

Figure 4. *Percentages of formal errors that are nouns as a function of (p − s) and chance expectations. Figure based on data from Schwartz, Dell, Martin, Gahl, and Sobel (2006).*

Schwartz et al. (2006) tested this prediction with their case series by selecting all formal errors and determining the extent to which they were nouns, and then binning those errors by the (p − s) values of the patients that generated them. Thus, a greater (p − s) should be associated with a greater tendency for lexical-level formals and hence a greater tendency for these to be nouns. The predicted trend was confirmed (Figure 4), thus supporting the dual nature of formals and the model's claims about processing steps and interaction. Case-series methods are critical to this finding. Without the common set of measures on the sample, it is not possible to fit the model to the patients to determine their parameters. And without a sizeable patient sample with variation in those parameters and in formal-error production, it is not possible to test for the trend relating properties of those errors to the parameters. Note, also, that this is clearly a case in which heterogeneity in the sample—here with regard to potential mechanisms for formal errors—is useful.

ANATOMICAL CASE SERIES

Traditional cn is considered to be part of cognitive psychology and hence is concerned with cognitive function rather than neural implementation (Caramazza & Coltheart, 2006; Coltheart, 2006). Over the last 10 years, though, the sharp division between cognitive psychology and neuroscience has been breaking down. We see evidence of this in computational models of behavioural data that build in anatomical and/or neurophysiological assumptions (Gotts & Plaut, 2002; Lambon Ralph, McClelland, Patterson, Galton, & Hodges, 2001; Plaut, 2002) and in brain imaging and lesion localization studies that address issues of relevance to cognitive theory (Bedny, Caramazza, Grossman, Pascual-Leone, & Saxe, 2008; Damasio, Grabowski, Tranel, Hichwa, & Damasio, 1996). A recent study makes this general point through a new application of the cn case series method that we call the *anatomical case series*.

The study used voxel-based lesion–symptom mapping (VLSM) to identify left hemisphere voxels that, when lesioned, are associated with semantic errors in naming (Schwartz et al., 2009). Behavioural data were obtained from 64 patients with chronic, poststroke aphasia, along with contemporaneous, high-quality computerized brain scans. For each patient, there were three behavioural scores: proportion of semantic errors in naming, verbal comprehension accuracy, and nonverbal comprehension accuracy. Additionally, two derived scores were computed, representing the semantic error score after controlling for verbal and nonverbal comprehension, respectively. The two derived scores index, in slightly different ways, semantic errors generated during the production stage of lexical access and not during the semantic stage (i.e., target conceptualization). After the lesions were registered to a common template, the association between patients' semantic error score and their lesion status (presence vs. absence of a lesion) was tested in each voxel using *t* tests. The results were clear: One area of the brain showed an association with semantic errors when comprehension was controlled (i.e., using the derived scores). That area was the left anterior temporal lobe (L ATL), especially mid to anterior middle temporal gyrus. A follow-up study showed that the association between semantic errors and L ATL lesions also survived correction

for phonological errors and for total lesion size (Walker et al., 2010).

Finding an area of the brain that is specifically associated with the production of semantic errors, above and beyond its association with comprehension ability and phonological error production, constitutes the anatomical footprint of the "postsemantic, prephonological" generation of semantic errors (Caramazza & Hillis, 1990; Dell et al., 1997; Graham et al., 1995; Rapp & Goldrick, 2000). Of course, this mechanism for semantic errors has been hypothesized in cognitive theories of language production. For example, it is a central premise of the interactive, two-step model and is readily compatible with any model that postulates a lexical level in between semantics and phonology (Levelt, Roelofs, & Meyer, 1999). The fact that the area in question happens to be in the left ATL carries other implications—for example, for debates about the precise role of the ATL and its anatomical subdivisions in semantics and naming (for discussion, see Walker et al., 2010).

This example of an anatomical study with cognitive ramifications brings to mind these prescient words from Tim Shallice, written more than 20 years ago:

> I have argued that the use of group studies is not likely to lead to rapid theoretical advance and that, in general, information about the localization of lesions is not vital for cognitive neuropsychology. However, this is not to argue that group studies and information on lesion localization should be excluded from cognitive neuropsychology. The rather negative assessment made of group studies and localisation information is not one of principle, but a pragmatic one specific to the methods available at present. With, say, advances in neurological measurement techniques, the situation might very well change. (Shallice, 1988, p. 214)

We believe that the anatomical case series is exactly the kind of "group study" with "localization information" that, because of advancements in neurological measurement, now warrants inclusion as a vital methodological tool in cognitive neuropsychology.

HETEROGENEITY IN CASE SERIES

As we mentioned earlier, the standard cn concern with group studies is the likelihood of heterogeneity of the deficits in the group, making the group mean, at best, unrepresentative of all cases and, at worst, irrelevant. Although case series are less concerned with means than with more complex trends, the lack of homogeneity of the groups can also limit conclusions there, because the trends themselves can be heterogeneous. But in a case series, one has the potential to understand this heterogeneity to advance theory. We consider two particular approaches. The first is what we call "track down the deviations". Because case series analysis creates some kind of model (even if it is only a simple linear regression), each patient's measurement can be characterized as consistent with the model or not, and the explanation for the deviating cases can be sought, by using single-subject style assessment. We already saw this approach taken in the Schwartz et al. (2006) study. Another example is Fischer-Baum and Rapp's (2008) case series study of perseveration in spelling. These authors examined perseverations of letters from previous trials in 12 dysgraphic patients. A failure-to-activate account of perseverations predicts that each patient's perseveration proportion should be predictable from their rate of other nonperseveratory errors. This relation was found. However, Fischer-Baum and Rapp then used a precise technique to identify outlier patients from the function predicting perseverations from other errors. There was one clear outlier who perseverated much more than would be expected. Thus, this patient cannot be explained by the failure-to-activate account. Fischer-Baum and Rapp then did additional testing of the outlier patient finding that he perseverated in many tasks—spelling, copy transcoding, naming, immediate serial recall, and even in the retention of visual shapes. They hypothesized that, in this case, the patient suffered from a general deficit in inhibition. Thus, by tracking down the deviate in their case series, Fischer-Baum and Rapp used the heterogeneity in their sample to support the idea that some perseverations are simply a manifestation of weak retrieval (the failure-to-activate account), but others result from a failure to a specific inhibitory mechanism.

A second way to confront sample heterogeneity is to attempt to remove it. The simplest method is to remove patients based on available measures. For example, Dell et al. (1997) eliminated patients who made many omission or articulatory errors from their sample, so as to focus on patients who make errors of commission that arise during lexical access. Another way to dispose of heterogeneity is statistical control, as, for example, when Schwartz et al. (2009) used regression to control for the effect of comprehension ability on the production of semantic errors. This leads to a residualized dependent variable, in this case, semantic errors made above and beyond what would be expected from any central semantic deficit that is present. Similarly, one can draw conceptual distinctions within one's dependent variable to create more uniform measures. So, instead of a raw measure of "semantic" errors in naming, Schwartz et al. (in press) separately analysed taxonomically related errors ("cat" for "dog") and thematically related errors ("bone" for "dog") and found that they had distinct neural correlates.

By using these methods, one can limit the heterogeneity of the patient sample and the data. Of course, as we have emphasized, one does not want to throw out the baby with the bath water. A case series depends on there being some differences in the measurements taken across patients. Fully homogenizing a sample just leaves the researcher with a collection of indistinguishable patients.

CASE SERIES AND SINGLE-SUBJECT METHODS

We maintain that single-subject and case series methods complement one another. Single-subject studies lead to the discovery of hypothesized cognitive mechanisms that are thought to interact in particular ways. Often these interactions can be expressed as predictions about quantitative trends regarding behavioural covariation, and questions can be posed regarding covariation between cognitive mechanisms and lesion locations. If so, a case series can provide a test. And, after that, patients that deviate from the principal trends can be examined using single-subject techniques to track down the source of the deviation.

Ultimately, both single-subject and case series methods are, as they should be, centred on explaining variability; it is just that the sources of variability differ. In a single-case study, the variability comes from the many tests that are administered to the patient. Some tests reveal deficits (to varying degrees), and others do not. This variation allows the researcher to draw conclusions. The same applies to a case series, except that there are more cases, but typically fewer tests. The variability in patient performance on the tests and, particularly, the covariation between tests, provide the basis for scientific inference.

Without doubt, case series and single-subject studies can be in tension, as when a theoretically meaningful association, established through large case series investigations, is violated by compelling evidence of functional dissociation in one or a few cases. Yet as long as one considers this the beginning of the story and not the end, considerable progress can be made in explaining why the association is partial rather than complete (e.g., Dilkina et al., 2008) or, from the alternative perspective, how a dual-mechanism, dissociable system manages to associate so often (Coltheart et al., 2001). One thing is clear: The growing sophistication of statistical techniques of lesion analysis that control for functional and anatomical confounds (Schwartz et al., 2009) provides a unique opportunity: The possibility that a strong partial association is merely an accident of anatomical overlap—a classic argument for why associations data are suspect in cn (e.g., Marin, Saffran, & Schwartz, 1976; Shallice, 1988)—can now actually be tested with anatomical data.

CONCLUDING THOUGHTS

The preceding review and analysis of case series methodology have described its many strengths. To summarize, case series are good for:

- Identifying quantitative trends.
- Revealing curvilinear, particularly nonmonotonic relations.
- Jointly considering associations and dissociations, where a large enough *N* affords statistical power to detect complex associations and the potential to identify theoretically important dissociations.
- Providing data for modelling, particularly modelling of how patient scores vary with severity and other individual difference variables.
- Providing data for advanced lesion analyses and thereby interfacing with theories of both brain and cognition.

The review has also identified certain weaknesses or issues that require the researcher to take care:

- There is the temptation to stop with a statistically significant correlation and not track down deviating cases, theoretically important dissociations, or nonlinear relations.
- There is the danger of committing prematurely to the functional significance of a demonstrated association, without carefully considering alternative possibilities, particularly alternatives that cannot easily be assessed with the limited set of measures taken. For anatomical case series, uninteresting explanations for an association such as lesion size or overlap must be ruled out.
- There are no rules for how broadly or narrowly to set the inclusion criteria for the sample, and misjudgements in either direction can have significant consequences—for example, limiting the range with too homogeneous a group, or obscuring important relationships with one that is too heterogeneous.
- The resources required to conduct a well-designed cn-style case series, with ample sample size and adequate background testing, are formidable. Consequently, the approach may be possible only for a few institutions or groups.

While most of the identified weaknesses are avoidable and, with the usual back-and-forth between advocates and critics, self-correcting, the last one is more troubling. To address this problem, we believe that individuals and organizations should promote greater sharing of patients (within the bounds of geography and confidentiality), test materials, and data. As a step in this direction a large, searchable web-based database available to the research community was developed and is described in an accompanying paper (Mirman et al., 2011).

In closing, we return to something we asserted at the beginning—namely, that case series methodology should, indeed will, continue to be an important part of cn. Future researchers who study cognitive impairments in patients will work collaboratively to amass data that are treated by both single-subject and multiple-subjects methods. They will discover associations and dissociations, and these discoveries will have relevance not just for cognitive psychology, but for all scientific and clinical disciplines with which we share interests.

Manuscript received 23 August 2010
Revised manuscript received 6 March 2011
Revised manuscript accepted 7 March 2011
First published online 5 July 2011

REFERENCES

Adlam, A. L., Patterson, K., Rogers, T. T., Nestor, P. J., Salmond, C. H., Acosta-Cabronero, J., et al. (2006). Semantic dementia and fluent primary progressive aphasia: Two sides of the same coin? *Brain*, *129*(11), 3066–3080.

Bedny, M., Caramazza, A., Grossman, E., Pascual-Leone, A., & Saxe, R. (2008). Concepts are more than percepts: The case of action verbs. *Journal of Neuroscience*, *28*(44), 11347–11353.

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

Bozeat, S., Lambon Ralph, M. A., Patterson, K., Garrard, P., & Hodges, J. R. (2000). Non-verbal semantic impairment in semantic dementia. *Neuropsychologia*, *38,* 1207–1215.

Caramazza, A., & Coltheart, M. (2006). *Cognitive Neuropsychology* twenty years on. Cognitive Neuropsychology, *23*(1), 3–12.

Caramazza, A., & Hillis, A. E. (1990). Where do semantic errors come from? *Cortex*, *26*(1), 95–122.

Cipolotti, L., & Warrington, E. K. (1995). Semantic memory and reading abilities: A case report. *Journal of the International Neuropsychology Society*, *1,* 104–110.

Coltheart, M. (2006). Acquired dyslexias and the computational modelling of reading. *Cognitive Neuropsychology*, *23*(1), 96–109.

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

Damasio, H., Grabowski, T. J., Tranel, D., Hichwa, R. D., & Damasio, A. (1996). A neural basis for lexical retrieval. *Nature*, *380*(6574), 499–505.

Dell, G. S., & Caramazza, A. (2008). Introduction to special issue on computational modeling in cognitive neuropsychology. *Cognitive Neuropsychology*, *25,* 131–135.

Dell, G. S., Lawler, E. N., Harris, H. D., & Gordon, J. K. (2004). Models of errors of omission in aphasic naming. *Cognitive Neuropsychology*, *21*(2–4), 125–145.

Dell, G. S., Martin, N., & Schwartz, M. F. (2007). A case series test of the interactive two-step model of lexical access: Predicting word repetition from picture naming. *Journal of Memory and Language*, *56,* 490–520.

Dell, G. S., & O'Seaghdha, P. G. (1992). Stages of lexical access in language production. *Cognition*, *42*(1–3), 287–314.

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review*, *104,* 801–838.

Dilkina, K., McClelland, J. L., & Plaut, D. C. (2008). A single-system account of semantic and lexical deficits in five semantic dementia patients. *Cognitive Neuropsychology*, *25*(2), 136–164.

Farrington, C. P., Nash, J., & Miller, E. (1996). Case series analysis of adverse reactions to vaccines: A comparative evaluation. *American Journal of Epidemiology*, *143*(11), 1165–1173.

Fischer-Baum, S., & Rapp, B. (2008, October). *Underlying causes of letter perseveration errors in dysgraphia*. Talk given at the 45th Annual Meeting of the Academy of Aphasia, Turku, Finland.

Freud, S. (1953). *On aphasia*. (E. Stengel, Trans.). New York City, NY: International University Press (Original work published 1891).

Garrard, P., Lambon Ralph, M. A., Hodges, J. R., & Patterson, K. (2001). Prototypicality, distinctiveness, and intercorrelation: Analyses of the semantic attributes of living and nonliving concepts. *Cognitive Neuropsychology*, *18*(2), 125–174.

Goldrick, M. (2008). Does like attract like? Exploring the relationship between errors and representational structure in connectionist networks. *Cognitive Neuropsychology*, *25,* 287–313.

Goldrick, M., & Rapp, B. (2002). A restricted interaction account (RIA) of spoken word production: The best of both worlds. *Aphasiology*, *16,* 20–55.

Gotts, S. J., & Plaut, D. (2002). The impact of synaptic depression following brain damage: A connectionist account of "access/refractory" and "degraded-store" semantic impairments. *Cognitive, Affective, and Behavioral Neuroscience*, *2,* 187–213.

Graham, K., Patterson, K., & Hodges, J. R. (1995). Progressive pure anomia: Insufficient activation of phonology by meaning. *Neurocase*, *1,* 25–38.

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

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59,* 434–446.

Jefferies, E., & Lambon Ralph, M. A. (2006). Semantic impairment in stroke aphasia versus semantic dementia: A case series comparison. *Brain*, *129,* 2132–2147.

Kittredge, A. K., Dell, G. S., Verkuilen, J., & Schwartz, M. F. (2008). Where is the effect of frequency in word production? Insights from aphasic picture naming errors. *Cognitive Neuropsychology*, *25*(4), 463–492.

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

Lambon Ralph, M. A., McClelland, J., Patterson, K., Galton, C., & Hodges, J. R. (2001). No right to speak? The relationship between object naming and semantic impairment: Neuropsychological evidence and a computational model. *Journal of Cognitive Neuroscience*, *13,* 341–356.

Lambon Ralph, M. A., Sage, K., Jones, R. W., & Mayberry, E. J. (2010). Coherent concepts are computed in the anterior temporal lobes. *Proceedings of the National Academy of Sciences*, *107*(6), 2717–2722.

Levelt, W. J.M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22,* 1–75.

Marin, O. S. M., Saffran, E. M., & Schwartz, M. F. (1976). Dissociation of language in aphasia: Implications for normal function. *Annals of the New York Academy of Sciences*, *280,* 868–884.

Martin, N., Dell, G. S., Saffran, E. M., & Schwartz, M. F. (1994). Origins of paraphasias in deep dysphasia: Testing the consequences of a decay impairment to an interactive spreading activation model of lexical retrieval. *Brain and Language*, *47,* 609–660.

McCloskey, M. (1993). Theory and evidence in cognitive neuropsychology: A "radical' response to Robertson, Knight, Rafal, and Shimamura (1993). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 718–734.

McCloskey, M., & Caramazza, A. (1988). Theory and methodology in cognitive neuropsychology: A response to our critics. *Cognitive Neuropsychology*, 5, 583–623.

McRae, K., de Sa, V. R., & Seidenberg, M. S. (1997). On the nature and scope of featural representations for word meaning. *Journal of Experimental Psychology: General*, *126,* 99–130.

Mirman, D., Strauss, T., Brecher, A., Walker, G., Sobel, P., Dell, G., et al. (2011). *A large, searchable, web-based database of aphasic performance on picture naming and other tests of cognitive function*. Cognitive Neuropsychology.

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

Mummery, C. J., Patterson, K., Wise, R. J. S., Vandenbergh, R., Price, C. J., & Hodges, J. R. (1999). Disrupted temporal lobe connections in semantic dementia. *Brain*, *122,* 61–73.

Nickels, L. (1995). Getting it right? Using aphasic naming errors to evaluate theoretical models of spoken word recognition. *Language and Cognitive Processes*, *10,* 13–45.

Nickels, L., & Howard, D. (1994). A frequent occurrence? Factors affecting the production of semantic errors in aphasic naming. *Cognitive Neuropsychology*, *11,* 289–320.

Nickels, L., & Howard, D. (1995). Phonological errors in aphasic naming: Comprehension, monitoring and lexicality. *Cortex*, *31,* 209–237.

Nickels, L., & Howard, D. (2004). Dissociating effects of number of phonemes, number of syllables, and syllabic complexity on word production in aphasia: It's the number of phonemes that counts. *Cognitive Neuropsychology*, *21,* 57–78.

Nozari, N., Kittredge, A. K., Dell, G. S., & Schwartz, M. F. (2010). Naming and repetition in aphasia: Steps, routes, and frequency effects. *Journal of Memory and Language*, 63, 541–559.

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

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*(2), 169–183.

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, *8,* 976–987.

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

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,* 767–808.

Plaut, D. C. (2002). Graded modality-specific specialisation in semantics: A computational account of optic aphasia. *Cognitive Neuropsychology*, *19*(7), 603–639.

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*(1), 56–115.

Plaut, D. C., & Shallice, T. (1993). Deep dyslexia: A case study of connectionist neuropsychology. *Cognitive Neuropsychology*, *10,* 377–500.

Rapp, B. C., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review*, *107*(3), 460–499.

Rogers, T. T., Hocking, J., Noppeney, U., Mechelli, A., Gorno-Tempini, M. L., Patterson, K., et al. (2006). Anterior temporal cortex and semantic memory: Reconciling findings from neuropsychology and functional imaging. *Cognitive, Affective, and Behavioral Neuroscience*, *6*(3), 201–213.

Rogers, T. T., Lambon Ralph, M. A., Garrard, P., Bozeat, S., McClelland, J. L., Hodges, J. R., et al. (2004). Structure and deterioration of semantic memory: A neuropsychological and computational investigation. *Psychological Review*, *111*(1), 205–235.

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*(2), 331–352.

Ruml, W., & Caramazza, A. (2000). An evaluation of a computational model of lexical access: Comment on Dell et al. (1997). *Psychological Review*, *107*(3), 609–634.

Ruml, W., Caramazza, A., Capasso, R., & Miceli, G. (2005). Interactivity and continuity in normal and aphasic language production. *Cognitive Neuropsychology*, *22,* 131–168.

Schwartz, M. F., Dell, G. S., Martin, N., Gahl, S., & Sobel, P. (2006). A case series test of the interactive two-step model of lexical access: Evidence from picture naming. *Journal of Memory and Language*, *54,* 228–264.

Schwartz, M. F., Kimberg, D. Y., Walker, G. M., Brecher, A., Faseyitan, O., Dell, G. S., et al. (in press). *A neuroanatomical dissociation for taxonomic and thematic knowledge in the human brain. Proceedings of the National Academy of Sciences.*

Schwartz, M. F., Kimberg, D. Y., Walker, G. M., Faseyitan, O., Brecher, A., Dell, G. S., et al. (2009). Anterior temporal involvement in semantic word retrieval: VLSM evidence from aphasia. *Brain*, *132*(12), 3411–3427.

Schwartz, M. F., Saffran, E. M., & Marin, O. S. M. (1980). Fractionating the reading process in dementia: Evidence for word-specific print-to-sound associations. M. Coltheart, K. Patterson, & J.C. Marshall (Eds.), *Deep dyslexia*. London, UK: Routledge & Kegan Paul.

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

Walker, G. M., Schwartz, M. F., Kimberg, D. Y., Faseyitan, O., Brecher, A., Dell, G. S., et al. (2010). Support for anterior temporal involvement in semantic error production in aphasia: New evidence from VLSM. *Brain and Language*. Advance online publication. doi:10.1016/j.bandl.2010.09.008

Warrington, E. K. (1975). The selective impairment of semantic memory. *Quarterly Journal of Experimental Psychology*, *27,* 635–657.

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

Woollams, A., Lambon Ralph, M. A., Plaut, D., & Patterson, K. (2007). SD-squared: On the association between semantic dementia and surface dyslexia. *Psychological Review*, *114*(2), 316–339.