Available online at www.sciencedirect.com
ScienceDirect
Journal of Memory and Language 59 (2008) 524–544
Journal of Memory and Language
www.elsevier.com/locate/jml

Methods of testing and diagnosing model error: Dual and single route cascaded models of reading aloud

James S. Adelman *, Gordon D.A. Brown

Department of Psychology, University of Warwick, Gibbet Hill Road, Coventry, CV4 7AL, UK

Received 1 March 2007; revision received 2 October 2007
Available online 14 January 2008

Abstract

Models of visual word recognition have been assessed by both factorial and regression approaches. Factorial approaches tend to provide a relatively weak test of models, and regression approaches give little indication of the sources of models' mispredictions, especially when parameters are not optimal. A new alternative method, involving regression on model error, combines these two approaches with parameter optimization. The method is illustrated with respect to the dual route cascaded model of reading aloud. In contrast to previous investigations, this method provides clear evidence that there are parameter-independent problems with the model, and identifies two specific sources of misprediction made by model.

Keywords: Model testing; Regression analyses; Dual route cascaded model; Word naming; Visual word recognition

The ideal of modeling cognitive processes involves testing the consequences of the processes conjectured in models against behavioral data. In the context of visual word recognition, where many of the relevant predictors are properties of words, two approaches have emerged to testing such models. The first is a *factorial approach* in which the question is whether a model can simulate some set of emprically observed significant effects. The second is a *regression approach* in which the regression performance (correlation with data) of the model's behavior is compared to a regression equation with the relevant predictors.

Coltheart, Rastle, Perry, Langdon, and Ziegler (2001) applied both approaches to the assessment of their dual route cascaded (DRC) model of reading aloud. This model incorporates a lexical (vocabulary-based) interactive-activation (IA) route and a nonlexical (spelling-sound rule-based) route that interact (collaborate or compete) to generate pronunciations. They argued, primarily on the basis of the results of the factorial approach, that "the DRC model is the most successful of current computational models of reading" (p. 251). They also argued that "if there is no other theory that ... [is] both complete and sufficient, resting on laurels is a reasonable thing to do until the emergence of a competitor." (p. 204). It is our contention that neither the factorial approach nor the regression approach nor their combination, as currently applied, is adequate to demonstrate that a model is sufficient to explain effects or data. One reason is that the parameters—numerical values defined to control the operation of a model without a fully specified relationship with independent variables—used in such analyses are not chosen to be optimal for the data set in question; in

* Corresponding author.
E-mail address: J.S.Adelman@warwick.ac.uk (J.S. Adelman).

0749-596X/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2007.11.008

consequence, the criterion for sufficiency is artificially low to compensate.

In this paper, we (i) set out the problems with the existing approaches; (ii) introduce a method of combining the two approaches with optimal parameter selection that resolves several of these problems; (iii) describe the DRC model and our optimization of its parameters for the Spieler and Balota (1997) data set, which was among those examined by Coltheart et al. (2001); (iv) demonstrate the theoretical leverage the approach can obtain with the application of the technique to the DRC; and (v) discuss issues surrounding the analyses, including the data collection strategies that support the method. We choose the DRC model for this exposition for three reasons. First, this model is a complete model of reading aloud that appears to be able to account for a relatively wide range of effects, indeed one that is wider than other models, and thus is in need of strict testing. Second, the model is relatively well-suited to the approach described here. Finally, Coltheart et al. have argued against using the optimization of parameters and in favor of a pure factorial approach, despite some examination of regression approach results.

The factorial approach

The *factorial approach* to deciding whether a model is adequate to explain a particular cognitive process is to generate a list of *effects* (including interactions) that one has reason to believe occur in the task of interest, and then determine whether the model can predict the direction of these effects correctly.

Coltheart et al. (2001), for instance, list 18 effects in word naming for which their model can account, and of these, six relate to a standard word naming task (i.e., involve neither nonwords nor priming): (i) regularity, (ii) frequency, (iii) their interaction, (iv) position of (first) irregularity, (v) rime consistency, and (vi) length. One further effect in standard word naming that Coltheart et al. attempt to account for with their model is the orthographic neighborhood size (*N*; Andrews, 1989, 1992; Coltheart, Davelaar, Jonasson, & Besner, 1977) effect. They find that the model with their standard parameters does not predict an effect of orthographic *N*, but that an alteration to one of the parameters causes the model to predict such an effect. Thus it was demonstrated that the DRC *can* predict an orthographic *N* effect.

Suboptimal parameter sets

This change in parameters to account for orthographic *N* effects raises the first problematic issue for the factorial approach: Whether and when it is appropriate to simulate each data set with different parameters. Coltheart et al.'s (2001) assert that they "would in any case not be interested in an approach in which each set of human data is simulated with a different set of DRC parameters" (p. 218). However, the DRC model is supposed (through parameter changes) to account for (i) strategic effects that are caused by changes in the stimulus list (see also Rastle & Coltheart, 1999) and (ii) individual differences (arising from the strengths of the different routes). Since each set of human data presumably comes from a different experiment with different items and participants, it is hard to see why the parameters should not be different. Indeed, we would argue that if one is following the good practice of seeking to disconfirm proposed theories, it is necessary to optimize parameters in order to avoid the contention that the chosen parameters are at fault, rather than the mechanisms of the model under test.

Incompleteness of data

A corollary of the specificity of parameters to experiments is that the results of any given experiment speak only to the existence of some pattern under the specific conditions (and hence parameters) of that experiment, and not to those of another experiment. For instance, Coltheart et al. (2001) suggest that "it will be necessary to study whether the DRC model can yield a beneficial effect of *N* on word naming through a modification of the parameter set that does not compromise its successful simulation of the other effects" (pp. 224–225). This suggestion requires the implicit assumption that *N* and position of irregularity effects will occur together in one experiment, and their model should account for such a pattern. However, whether such an account is in fact desirable depends on a number of empirical questions because these effects have previously been demonstrated in different experiments, with different items and different participants. This means the effects could be compatible with two different parameterizations of the correct model. That is, in the absence of a data set demonstrating the combined pattern, it might yet be the case that these effects are mutually exclusive, and a model could be correct to predict them only with different parameter sets. Whilst this particular example could be resolved with an experiment with both position of irregularity and *N* as factors, any single such experiment would not solve the problem for other pairs of predictors, nor triples, quadruples and so on. Factorial designs with few factors will thus lead to ambiguities as to whether sets of effects should co-occur with the same parameters. Attempts to include more factors will include increasingly many items if the experiments are not to lose all their power. These items will necessarily be less well controlled (see *Item Selection* below), the analysis will tend therefore to rely on statistical partialing (i.e., regression)—and note that has already been the case for at least one study of the