

When parallel processing in visual word recognition is not enough: New evidence from naming

MARTHA ANNE ROBERTS

University of Waterloo, Waterloo, Ontario, Canada

KATHLEEN RASTLE and MAX COLTHEART

Macquarie University, Sydney, New South Wales, Australia

and

DEREK BESNER

University of Waterloo, Waterloo, Ontario, Canada

Low-frequency irregular words are named more slowly and are more error prone than low-frequency regular words (the regularity effect). Rastle and Coltheart (1999) reported that this irregularity cost is modulated by the serial position of the irregular grapheme–phoneme correspondence, such that words with early irregularities exhibit a larger cost than words with late ones. They argued that these data implicate rule-based serial processing, and they also reported a successful simulation with a model that has a rule-based serial component—the DRC model of reading aloud (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001). However, Zorzi (2000) also simulated these data with a model that operates solely in parallel. Furthermore, Kwantes and Mewhort (1999) simulated these data with a serial processing model that has no rules for converting orthography to phonology. The human data reported by Rastle and Coltheart therefore neither require a serial processing account, nor successfully discriminate among a number of computational models of reading aloud. New data are presented wherein an interaction between the effects of regularity and serial position of irregularity is again reported for human readers. The DRC model simulated this interaction; no other implemented computational model does so. The present results are thus consistent with rule-based serial processing in reading aloud.

Computational modeling has launched a new approach to studying cognition, wherein any theory about the nature of human information processing can be expressed in the form of a computer program that carries out that processing. In developing a computational model, it is necessary to specify exactly how information processing is carried out at each stage. As a result, such modeling forces exact specification of the cognitive theories to be modeled. Computational models have thus become important tools for detecting ambiguity and imprecision in cognitive theories. Once implemented, the model can be evaluated directly in order to test whether it provides

a sufficient account of human performance. Such evaluation often leads to the development of more refined theories of information processing.

Implemented computational models of reading aloud include the parallel distributed processing (PDP) model (Plaut, McClelland, Seidenberg, & Patterson, 1996), the connectionist dual process (CDP) model (Zorzi, Houghton, & Butterworth, 1998), the LEX model (Kwantes & Mewhort, 1999), and the dual route cascaded (DRC) model (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001). With several competing models on the table, a program of study devoted to adjudicating among the contenders is necessary in order for one to achieve a better understanding of human word recognition processes.

All of the models above posit different procedures for computing phonology from print. The DRC and LEX models each propose that at least one of these procedures operates serially from left to right across the letter string, whereas the PDP and CDP models are parallel processing models. Furthermore, the LEX and DRC models differ in fundamental ways that may also offer grounds for discriminating between them. Specifically, the LEX model does not implement rule-based grapheme–phoneme conversion, whereas DRC's nonlexical route does. The cur-

This work was supported by the Natural Sciences and Engineering Research Council of Canada Grant AO998 to D.B. and was submitted to the University of Waterloo in partial fulfillment of the requirements of a master's degree for M.A.R. We are particularly grateful to Marco Zorzi for providing the consistency analysis reported here and for helpful discussion. We are also grateful to Peter Kwantes for providing the LEX model simulation data reported here and to Ken Forster, Chris Kello, Ken Paap, and an anonymous reviewer for their helpful comments and comprehensive reviews. Address correspondence to M. A. Roberts or D. Besner, Department of Psychology, University of Waterloo, Waterloo, ON, N2L 3G1 Canada (e-mail: ma3rober@watarts.uwaterloo.ca or dbesner@watarts.uwaterloo.ca).

rent research investigates the regularity effect and its interaction with serial position in reading aloud in order to discriminate among the models.

We begin with a brief introduction of the contending computational models of word recognition mentioned above. Specific details about how each of the models operates can be found in various publications (Coltheart et al., 2001; Kwantes & Mewhort, 1999; Plaut et al., 1996; Zorzi, 2000; Zorzi et al., 1998). For present purposes, only those details of each model's architecture that are relevant to the present research are briefly noted. Serial processing in reading aloud is then discussed, followed by a description of several studies leading to the present research.

THE MODELS

The Parallel Distributed Processing Model (Plaut et al., 1996)

The PDP model posits two routes for translating orthography to phonology. One route (as yet unimplemented) maps orthography to phonology via meaning. The implemented route consists of orthographic input units connected to a set of hidden units that are in turn connected to a set of phonological output units.

This model does not contain explicit word representations, nor does it have explicit rules for converting orthography to phonology. Through training, the model learns spelling-sound associations based on the degree to which the spelling-sound associations in a particular word are consistent with those of other words. The emphasis is on the degree to which the mappings among orthography, phonology, and semantics of a given input are consistent with those of other inputs. Processing a word within the fully trained model is based on learned connection weights resulting from a model architecture of parallel processing between orthographic, phonological, and semantic layers. In the fully trained model, information is represented in distributed patterns of activation over groups of processing units. In processing input, activation flows freely among the model's processing layers until the entire network settles on a stable pattern of activation corresponding to its interpretation of the input.

The Connectionist Dual-Process Model (Zorzi et al., 1998)

The CDP model has two nonsemantic routes for converting orthography to phonology, a sublexical assembly route and a lexical route. The sublexical route learns typical spelling-sound correspondences via a two-layer network. The lexical route converts print to speech via direct activation of the word's phonological representation, learned within a three-layer network that consists of orthographic input, hidden units, and phonological output. Thus, through learning, orthographic representations correspond to phonologic representations in a one-to-one fashion. The model does not implement an orthographic lexicon, so print directly activates the word's phonological representation. There is potential for inter-

action between the two routes in computing output, because there are two sources of information feeding into output.

Analogous to the PDP model, the fully trained CDP model converts orthography to phonology on the basis of mappings learned through exposure to a large corpus of words. Orthographic input representations are fully connected to phonological output representations. The important point about this model is that processing within both routes of the model is parallel in nature.

The LEX Model (Kwantes & Mewhort, 1999)

The LEX model is fundamentally different from all of the other models in that it consists of only a single routine. LEX is a model of retrieval from semantic memory.¹ Complex patterns in performance are assumed to reflect simple retrieval operations from a lexicon of 105,000 words. There are three major components to LEX: letter identification, a retrieval process, and response generation. To name a word, the system identifies a sequence of letters in parallel and stores them as an ordered set that is then used to guide retrieval processes. Retrieval of orthography is serial. Phonology and orthography are retrieved at the same time, and phonology is later cleaned up in a filtering process from the noise created by phonological activation produced in the process of identifying the correct orthography. The first letter is fed into the system, and all items in the lexicon that begin with that spelling are activated. The phonology of all items in the lexicon that begin with that letter are also activated. The second letter is then fed in, and all items in the lexicon that contain the first and second letters are activated. This procedure continues until the best possible match for the input letter string is located in the lexicon. At the same time as it retrieves spelling from the lexicon, the model retrieves all phonemic information associated with that particular spelling. After the model has settled on an orthographic representation of the input string, the phonemic information obtained in the process of retrieving orthography is filtered and the best possible phonemic representation is settled on. Response times represent the number of samples required in order to retrieve the data from the lexicon added to the time that it takes to filter out the best possible phonemic representation.

Thus the LEX model first identifies letters and then uses those letters to guide lexical retrieval. Once the string is activated in memory, the phonology gets activated along with it. The critical point is that there are no rules for generating phonology from print. A word's phonology is stored with a lexical representation and activated at the lexical retrieval stage. It is then filtered in the phonological buffer from phonological activations that were initially retrieved but are irrelevant.

The Dual-Route Cascaded Model (Coltheart et al., 2001)

The DRC model has two implemented nonsemantic routes for converting orthography to phonology: a nonlexical route and a lexical route. The nonlexical route

translates graphemes to phonemes on the basis of a set of grapheme–phoneme conversion rules. The important point about the DRC model is that letter identification is carried out in parallel, but subsequent grapheme–phoneme conversion occurs serially, letter by letter, from left to right. The lexical route utilizes a dictionary lookup procedure for converting orthography to phonology.

Summary

All of the models posit different mechanisms for converting orthography to phonology. The PDP and the CDP models are purely parallel processing models, in that the fully trained model architectures do not contain any serial processing components. The LEX model is a serial processing model that has no rules for converting orthography to phonology. Only the DRC model implements a nonlexical, serial processing route that has specific rules for converting orthography to phonology. Therefore, studies of human word reading may yield evidence relevant to discriminating among the serial and parallel processing accounts of reading aloud, as well as between the rule-based and non-rule-based serial processing models.

The Regularity Effect

Many investigators of visual word recognition make a theoretical distinction between words with typical grapheme–phoneme correspondences (“regular” words, such as WAVE, SAVE, GAVE) and those with atypical grapheme–phoneme correspondences (“exception” words, such as HAVE). The standard finding is that exception words are named more slowly than regular words, but typically only when words are low in frequency (e.g., PINT is slower to name than MINT). This finding is commonly referred to as the regularity, or exception, effect. The regularity effect is a benchmark against which different theoretical accounts of word recognition have been assessed (e.g., Paap & Noel, 1991; Plaut et al., 1996; Rastle & Coltheart, 1999; Zorzi, 2000).

The PDP model treats regular and exception words as existing along a continuum of consistency. Thus, this model accounts for the regularity effect because the learned connection weights between orthography and phonology are weaker for low-frequency exception words (which are highly inconsistent) than they are for low-frequency regular words (which are highly consistent). Consistency is broadly defined as the extent to which the mappings between the orthography and phonology of a particular word are consistent with those of other words. Thus, it takes longer for the system to settle on a correct output for low-frequency exception words than for low-frequency regular words.

The LEX model asserts that the source of the regularity effect for low-frequency words is in the phonological buffer. Irregular words share letter combinations with other words, but the phonemes associated with certain letters are not typical. Regular words share letter combinations with other words, and the phonemes associated with those letters are typical. Thus, the clarity of irregu-

lar words is less than that for regular words in the phonological buffer, so LEX needs more time to filter out a correct pronunciation for an irregular word.

Both the DRC and the CDP models assume that processing regular and exception words through the nonlexical and lexical routes results in different outcomes. If a word is regular, both routes produce the same output. If a word is irregular, the two routes will produce conflicting output because the lexical route produces the correct pronunciation while the nonlexical route produces an erroneous but regular pronunciation. To resolve this conflict takes time; low-frequency exception words are therefore named more slowly than low-frequency regular words.

The Regularity \times Serial Position Interaction

Coltheart and Rastle (1994) argued that if processing through the nonlexical route is serial, as is modeled in DRC, then the size of the regularity effect will depend on the irregular component’s location in the word. If the irregular component occurs early in the word (i.e., *chef* as opposed to *chief*), then the observed regularity effect would be larger than it would if the irregular component occurred later in the word (i.e., *swap* as opposed to *snapp*). Early in the processing of words such as *chef*, the nonlexical route and the lexical route produce conflicting output. It takes time to resolve the conflict, thus resulting in a large regularity effect. However, in words such as *swap*, processing through the parallel lexical route has been completed to the extent that conflicting output from the nonlexical route results in relatively little cost of irregularity. Coltheart and Rastle (1994) and Rastle and Coltheart (1999) found such a regularity \times serial position interaction in a naming experiment with human subjects, and Rastle and Coltheart successfully simulated this interaction with the DRC model. They concluded that “models which operate solely in parallel do not predict this serial effect” (p. 488), and that “together, the human and DRC data strongly suggest that a serial process is at work in reading aloud” (p. 492).

No regularity \times serial position interaction is observed when Rastle and Coltheart’s (1999) stimuli are named by the PDP model. However, both the CDP model and the LEX model successfully simulate this interaction (Kwantes & Mewhort, 1999; Zorzi, 2000). Thus, Rastle and Coltheart were premature in their conclusion that “models which operate solely in parallel do not predict this serial effect” and in asserting that the human results successfully adjudicated between contending computational models of reading aloud.

What Happened?

How could the CDP model, which consists only of parallel processing, reproduce a regularity \times serial position interaction that is claimed to arise because of an interaction between serial and parallel processing routines? It turns out that, in Rastle and Coltheart’s (1999) stimuli, serial position of irregularity was confounded

with grapheme–phoneme consistency in such a way that the most highly inconsistent items were also those in which the irregular component occurred early in the word; the CDP model is sensitive to this variable. Thus, the regularity × serial position interaction produced by the CDP model can be explained as a position-specific grapheme–phoneme consistency effect (Zorzi, 2000).

It is argued here that there is a better way of approaching the problem of model adjudication than the one taken by Rastle and Coltheart (1999). They found a set of stimuli for which humans produced a regularity × serial position interaction; then they tested the DRC model on these stimuli and found the same interaction. Because the DRC model is the only model that incorporates a rule-based serial processing route, they assumed that their data had successfully adjudicated among the models. Rastle and Coltheart’s error was to second-guess the models. They made a strong inference about how to discriminate among the models, but they failed to test all of the competitor models. A more powerful approach is to test all of the models *before* collecting human data to see whether the data discriminate among the different accounts. We therefore first examined the position of irregularity effect in each of four models, using a new set of stimuli. We then collected human data with the aim of adjudicating among the models.

EXPERIMENT 1

Simulations

Stimuli

A set of 52 low-frequency irregular words and 52 matched regular control words, listed in the Appendix, were selected from the CELEX English database (Baayen, Piepenbrock, & Van Rijn, 1993). All of these items were monosyllabic² words from three to seven letters long and had CELEX frequencies between 0 and 90. Thirty-four words were irregular in the second position and 18 were irregular in the third position. Control items were matched to the irregular words on initial phoneme, and were selected so that across items in each condition they were matched on the basis of word frequency {CELEX frequencies were not significantly different for regular and irregular words in Position 2 [36.8 vs. 37.6, $t(66) < 1$] or in Position 3 [40.6 vs. 43.9, $t(34) < 1$] and letter length {no difference between regular and irregular words in Position 2 [4.6 vs. 4.5, $t(66) < 1$] and Position 3 [5.0 vs. 4.9, $t(34) < 1$]}. (See notes 3 and 4.) No words irregular

in the first position were used, because once these selection criteria had been met there were not enough items to warrant their inclusion.

The grapheme–phoneme consistency of each of the items was determined by analyzing the output of the non-lexical component of the CDP model in the manner described by Zorzi (2000); items and their consistency values are contained in the Appendix.⁵ In contrast to the stimulus set of Rastle and Coltheart (1999), the results revealed no confounding between grapheme–phoneme consistency and serial position [$F(1,78) < 1$].

Stimuli were submitted to the PDP (Plaut et al., 1996), CDP (Zorzi et al., 1998), LEX (Kwantes & Mewhort, 1999), and DRC (Coltheart et al., 2001) models under their respective standard parameters for word naming. Items not contained in model training sets were excluded from simulations. Because of the resulting missing values, regularity was treated as a between-items factor in all of the simulation analyses.

Simulations

The PDP Model

Six irregular words and five control words were excluded from the test set because they do not exist in this model’s corpus. The model mispronounced one third-position irregular target (*swear*); this item was therefore excluded from the RT analysis. Mean number of processing cycles at each position and level of regularity are shown in Table 1.

Number of processing cycles per item was analyzed with an analysis of covariance in which regularity and serial position were between-items factors and neighborhood size and length were covariates. There was a main effect of regularity [$F(1,86) = 17.14$, $MS_e = 0.16$, $p < .001$], with irregular words taking more cycles to settle than did regular words. There was no main effect of position [$F(1,86) = 0.13$, $MS_e = 0.16$, $p > .7$], and no interaction between regularity and serial position [$F(1,86) = 0.10$, $MS_e = 0.16$, $p > .7$]. Thus, the PDP model does not produce a regularity × serial position interaction for these stimuli.

The CDP Model

Thirteen irregular and eight regular words were excluded from the stimulus set because they do not exist in this model’s corpus. The model mispronounced four second-position irregular words (*gauche*, *sieve*, *soot*, and

Table 1
Regularity as a Function of Serial Position (2nd or 3rd) for PDP, CDP, LEX, and DRC
Simulations (Number of Processing Cycles) and Reaction Time (RT, in Milliseconds),
With Percent Error, for Humans (Item Data)

	PDP		CDP		LEX		DRC		Humans (by Items)			
	2nd	3rd	2nd	3rd	2nd	3rd	2nd	3rd	2nd		3rd	
									RT	%E	RT	%E
Irregular	2.1	2.1	4.3	4.7	602	679	89.9	79.9	553	20.2	555	9.7
Regular	1.7	1.8	2.7	2.9	484	498	78.3	77.7	526	4.9	551	4.8
Difference	0.4	0.3	1.6	1.8	118	181	11.6	2.2	27	15.1	4	4.9

waltz); these items were excluded from the analysis. Mean number of processing cycles at each position and level of regularity are shown in Table 1.

Number of processing cycles per item was analyzed with an analysis of covariance in which regularity and serial position were factors and neighborhood size and length were covariates. There was a main effect of regularity [$F(1,73) = 60.47$, $MS_e = 0.95$, $p < .001$], with irregular words taking more cycles to settle than regular words. There was no main effect of position [$F(1,73) = 1.71$, $MS_e = 0.95$, $p > .19$], and no interaction between regularity and serial position [$F(1,73) = 0.30$, $MS_e = 0.95$, $p > .5$]. Thus, the CDP model does not produce a regularity \times serial position interaction for these items.

The LEX Model

All 104 stimuli were submitted to LEX for simulation. The model mispronounced three second-position irregular words (*chaff*, *comb*, *learnt*), two second-position regular words (*cheer*, *whiff*), one third-position irregular word (*dreamt*), and one third-position regular word (*cleft*). These items were excluded from the analysis. Mean number of processing cycles at each position and level of regularity are shown in Table 1.

Number of processing cycles per item was analyzed with an analysis of covariance in which regularity and serial position were factors and neighborhood size and length were covariates.⁶ There was a main effect of regularity [$F(1,90) = 10.72$, $MS_e = 4.43 \times 10^4$, $p < .005$], with irregular words taking more cycles to settle than regular words. There was no main effect of position [$F(1,90) = 0.94$, $MS_e = 4.43 \times 10^4$, $p > .3$], and no interaction between regularity and serial position [$F(1,90) = 0.49$, $MS_e = 4.88 \times 10^4$, $p > .4$]. Thus, the LEX model does not produce a regularity \times serial position interaction for these items.

The DRC Model

All 104 stimuli were submitted to the DRC model for simulation.⁷ The DRC model made no naming errors to these stimuli. Mean number of processing cycles at each position and level of regularity are shown in Table 1.

Number of processing cycles per item was analyzed with an analysis of covariance in which regularity and serial position were factors and neighborhood size and length were covariates. There was a main effect of regularity [$F(1,98) = 64.25$, $MS_e = 16.96$, $p < .001$], a main effect of position [$F(1,98) = 36.19$, $MS_e = 16.96$, $p < .001$], and an interaction between regularity and serial position [$F(1,98) = 30.00$, $MS_e = 16.96$, $p < .001$], in which a larger cost of regularity is observed for second-position items than for third-position items. Planned comparisons showed that the regularity effect was significant for both second-position items [$F(1,64) = 91.67$, $MS_e = 24.18$, $p < .001$] and third-position items [$F(1,32) = 12.72$, $MS_e = 3.00$, $p < .001$].

The DRC model thus produces the predicted regularity \times serial position interaction with this stimulus set;

the other models do not. An experiment that determined whether or not human reading data exhibit this interaction would therefore falsify at least one, and possibly three, of these four models. We turn now to the human data.

Human Data

Method

Subjects. The subjects were 30 undergraduate students from the University of Waterloo. All individuals had normal or corrected-to-normal vision and were native English speakers. Each individual was paid \$6.00 for his/her participation.

Materials. All 104 stimuli were used.

Apparatus and Procedure. Stimuli were displayed on a 15-in. ADI Micro Scan color monitor, using standard Micro Experimental Laboratory 2 (MEL 2) software (Schneider, 1988) font, in white on a black background. Stimulus presentation and data recording were controlled by MEL 2 software, which was run on a Vault Pentium II personal computer. Responses were collected using a voice-key and response latencies were measured to the nearest millisecond.

Subjects were tested individually and were seated approximately 50 cm from the computer monitor. They were instructed to read each item out loud as quickly and accurately as possible. Each trial began with a fixation cross displayed in the middle of the screen for 750 msec, after which time the cross was replaced with the item to be pronounced. The display terminated when a pronunciation was made. The experimenter then coded the pronunciation as correct, incorrect (regularization or other), or mistrial (cough, stutter). The subjects pressed a button to initiate the beginning of each trial. Each experimental session began with 16 practice trials, consisting of words not contained in the test set, followed by 104 experimental trials presented in random order.

Results

Responses classified as incorrect pronunciations or voice key failures were removed from the reaction time (RT) analysis, along with their matched controls. RTs for correct responses were then subjected to a recursive data trimming procedure in which the criterion cut-off for outlier removal is established by the sample size in each condition for each subject (Van Selst & Jolicœur, 1994). Outlier removal resulted in 1.4% of the data's being discarded. Of all pronunciation errors made, 98.7% were regularization errors. Mean correct RTs and percentage error for subjects at each level of regularity and serial position as analyzed by items are displayed in Table 1.

It should be noted that interpretation of the regularization data is somewhat ambiguous because some unknown proportion of these "errors" may have arisen simply because the subject did not know a particular word, rather than the nonlexical routine's having won out. With this reservation in mind, the error data are reported for the sake of completeness. We place more emphasis on the correct RT data, given that it reflects correct pronunciations. From the perspective of the dual route model, subjects cannot compute correct pronunciations for exception words that are unknown to them.

Subjects analysis. Mean correct RTs and percentage error for subjects were each analyzed with a 2×2 repeated measures analysis of variance with regularity (regular, irregular) and serial position of irregularity (Positions 2 and 3) as factors. Correct RTs yielded a main

effect of regularity [$F(1,29) = 4.52$, $MS_e = 618.69$, $p < .05$] and a main effect of position [$F(1,29) = 31.80$, $MS_e = 643.46$, $p < .001$]. More critically, the predicted interaction between regularity and serial position was observed [$F(1,29) = 20.15$, $MS_e = 154.41$, $p < .001$], such that a greater cost of irregularity was observed for second-position items than for third-position items. Planned comparisons revealed a significant effect of regularity for second-position items [$F(1,29) = 18.36$, $MS_e = 321.45$, $p < .001$], but not for third-position items [$F(1,29) = 0.01$, $MS_e = 451.65$, $p > .9$].

Analysis of the errors revealed a main effect of regularity [$F(1,29) = 112.82$, $MS_e = 0.29$, $p < .001$], a main effect of position [$F(1,29) = 29.10$, $MS_e = 0.34$, $p < .001$], and an interaction between regularity and serial position [$F(1,29) = 42.71$, $MS_e = 0.22$, $p < .001$] such that a greater cost of irregularity was observed for second-position items than for third-position items. Planned comparisons revealed an effect of regularity for second-position items [$F(1,29) = 111.28$, $MS_e = 0.35$, $p < .001$], as well as for third-position items [$F(1,29) = 22.17$, $MS_e = 0.16$, $p < .001$].

Items analysis. Correct RTs and percentage error for items were analyzed with an analysis of covariance with regularity (regular, irregular) and serial position of irregularity (Positions 2 vs. 3) as factors and neighborhood size and length as covariates. The standard problem with item analyses is that subject variance is incorporated into the item means, making it difficult to detect an effect (see Raaijmakers, Schrijnemakers, & Gremmen, 1999). Rather than not report item analyses, as Raaijmakers et al. suggest, we extracted subject variance by standardizing each subject's data.

This analysis yielded a marginal main effect of regularity [$F(1,98) = 2.54$, $MS_e = 0.32$, $p < .10$] and a main effect of position [$F(1,98) = 4.94$, $MS_e = 0.32$, $p < .05$]. The predicted interaction between regularity and serial position was marginal [$F(1,98) = 2.46$, $MS_e = 0.32$, $p < .10$]. Planned comparisons revealed a significant effect of regularity for second-position items [$F(1,64) = 6.43$, $MS_e = 0.34$, $p < .01$], but no effect of regularity for third-position items [$F(1,32) = 0.01$, $MS_e = 0.27$, $p > .9$].

Analysis of the item error data yielded a main effect of regularity [$F(1,98) = 14.04$, $MS_e = 170$, $p < .001$] and a marginally significant main effect of position [$F(1,98) = 3.60$, $MS_e = 170$, $p < .06$]. The interaction between regularity and serial position was marginal [$F(1,98) = 3.71$, $MS_e = 170$, $p < .06$]. Planned comparisons revealed a significant effect of regularity for second-position items [$F(1,64) = 18.04$, $MS_e = 211$, $p < .001$] and no effect of regularity for third-position items [$F(1,32) = 2.70$, $MS_e = 81$, $p > .1$].

DISCUSSION

The human data yielded an interaction between the effects of regularity and serial position of irregularity such that the cost of irregularity was larger for second-position items than for third-position items. This pattern was seen

in both the RT and the error data in the subject analysis, and in the standardized RT and error data in the item analysis, replicating the general findings of Coltheart and Rastle (1994) and Rastle and Coltheart (1999) with a better controlled stimulus set. In this instance, however, the implications for computational models are clear: Only the DRC model reproduced the pattern seen in the human data. The PDP, CDP, and LEX models all failed to produce an interaction between regularity and serial position. On the basis of these data, we suggest that the PDP, CDP, and LEX implementations are insufficient.

Now that we have established that only the DRC model can simulate the human data reported here, it is important that we consider whether the model simulates these findings for the correct reason. That is, it remains possible that human readers in this experiment and the DRC model show an interaction between regularity and serial position, but for different reasons. We argue that the DRC model provides a sufficient account of these findings; nevertheless we will evaluate three alternative accounts of the human data.

Can Effects of Onset Complexity Account for the Results?

Kawamoto and Kello (1999) have reported that words that begin with complex onsets (e.g., *slap*) are named significantly more quickly than words with simple onsets (e.g., *sap*; see also Rastle & Davis, 2002). Although the stimuli involved in the regularity comparison were matched for initial phoneme, we were not able to match for onset complexity. Complex onsets were present in none of the Position 2 irregular items, but they were present in 26.5% of regular items; and although complex onsets were present in 94% of Position 3 irregular items, they were present in only 83% of Position 3 regular items. Thus, effects of onset complexity might increase the size of the regularity effect at Position 2, but decrease it at Position 3.

We argue that although the effect of onset complexity may have contributed to the decrease in the size of the regularity effect across position of irregularity, it cannot account for the total effect that we observed. Consider that the onset complexity mismatch yields a net 26.5% of the Position 2 items that act to increase the size of the regularity effect, and a net 11% of the Position 3 items that act to decrease the size of the regularity effect. On the basis of the reported size of the onset complexity effect (14 msec, Kawamoto & Kello, 1999; 9 msec, Rastle & Davis, 2002), we would expect, at maximum, a 5-msec decrease in the size of the regularity effect over position of irregularity (37.5% * 14 msec)—far less than the 23-msec effect that we observed.

Can a Serial Response Execution System Account for the Results?

Kawamoto, Kello, Jones, and Bame (1998) argued that a simulation of the position of irregularity effect could be achieved by a model that operated solely in parallel if it were assumed that readers begin articulating as soon as the

initial phoneme is known, regardless of whether subsequent phonemes have been computed. Kawamoto et al. argued that in naming words with late irregularities (e.g., *crepe*, *swear*), articulation (and the emission of acoustic energy) begins before the irregular phoneme has been computed; therefore, no cost of irregularity is evident on naming latency. In contrast, in reading aloud words with early irregularities (e.g., *chef*, *bind*), the irregular phoneme must be computed before any acoustic energy is emitted; therefore a cost of irregularity is evident. Kawamoto et al. termed this theory the “initial phoneme criterion” for articulation and reported data consistent with this account. Rastle, Harrington, Coltheart, and Palethorpe (2000) argued that this theory is false on the basis of anticipatory coarticulation effects in speeded reading aloud. Since the position of the vocal organs at the onset of articulation was affected by the nature of the subsequent vowel phoneme, they argued that articulation in speeded reading aloud does not begin until vowel phonology has been computed. No subsequent account of the original data reported by Kawamoto et al. which deals effectively with anticipatory coarticulation has been developed.

Why Does the Position Manipulation Affect Regular Words?

Although the size of the regularity effect at Position 3 was 23 msec smaller than the size of the regularity effect at Position 2, the form of the interaction between regularity and position of irregularity merits comment. Specifically, the effect of serial position is manifest in the regular rather than the irregular words; position does not affect the latency with which irregular words are named. The form of the interaction in the DRC model is quite different, and this might mean that human readers and the DRC model show a decrease in the size of the regularity effect across serial position for different reasons.

We believe that discounting the DRC model in this instance would be hasty, however. Although stimuli across the regularity dimension were matched within position, Position 2 items were not matched to Position 3 items on several variables—most importantly, initial phoneme. The nature of a word’s initial phoneme is known to have a powerful effect on naming latency (see Treiman, Mullennix, Bijeljac-Babic, & Richmond-Welty, 1995), perhaps because the temporal relationship between articulation and acoustic energy may not be equivalent for all types of phoneme (Fowler, 1979), and because different types of phoneme may introduce unequal levels of voice key measurement error (see Rastle & Davis, 2002). It is possible that our inability to match across the position manipulation on this factor introduced a general slowing of response times—a 27-msec slowing—for Position 3 items (we are, of course, unable to offer a precise analysis of the probable effects of this variable, since there are no models of either the articulatory–acoustic temporal relationship across different types of phoneme, or the relationship between acoustic energy and voice key detection for different types of phoneme). At the same time,

Position 3 irregular words have a naming latency advantage relative to Position 2 irregular words—a later occurring irregularity—that cancels the slowing introduced by phonetic factors. The result is an interaction in which the regular words—not the irregular words—are slowed by the position manipulation. We would not expect the DRC model to produce this form of the interaction between regularity and position of irregularity, because the DRC model is not sensitive to the variance in RT introduced by initial phoneme.

Conclusions

We conclude that the human data reported here are inconsistent with the PDP, CDP, and LEX models, but are consistent with the DRC model of reading aloud. Furthermore, only the DRC model has a component that accomplishes nonlexical orthography–phonology translation serially with a rule-based translation mechanism. Thus, the human data are consistent with the claim that a rule-based serial mechanism contributes to the process of reading aloud.

REFERENCES

- BAAYEN, R. H., PIEPENBROCK, R., & VAN RIJN, H. (1993). *The CELEX lexical database* (CD-ROM). Philadelphia: University of Pennsylvania, Linguistic Data Consortium.
- COLTHEART, M., & RASTLE, K. (1994). Serial processing in reading aloud: Evidence for dual-route models of reading. *Journal of Experimental Psychology: Human Perception & Performance*, *20*, 1197-1211.
- 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.
- FOWLER, C. A. (1979). “Perceptual centers” in speech production and perception. *Perception & Psychophysics*, *25*, 375-388.
- KAWAMOTO, A. H., & KELLO, C. T. (1999). Effect of onset cluster complexity in speeded naming: A test of rule-based approaches. *Journal of Experimental Psychology: Human Perception & Performance*, *25*, 361-375.
- KAWAMOTO, A. H., KELLO, C. T., JONES, R., & BAME, K. (1998). Initial phoneme versus whole-word criterion to initiate pronunciation: Evidence based on response latency and initial phoneme duration. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *24*, 862-885.
- KWANTES, P. J., & MEWHORT, D. J. K. (1999). Modeling lexical decision and word naming as a retrieval process. *Canadian Journal of Experimental Psychology*, *53*, 306-315.
- PAAP, K. R., & NOEL, R. W. (1991). Dual route models of print to sound: Still a good horse race. *Psychological Research*, *53*, 13-24.
- PLAUT, D. C., MCCLELLAND, J. L., SEIDENBERG, M. S., & PATTERSON, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, *103*, 56-115.
- RAAIJMAKERS, J. G. W., SCHRIJNEMAKERS, J. M. C., & GREMMEN, F. (1999). How to deal with “the language-as-fixed-effect fallacy”: Common misconceptions and alternative solutions. *Journal of Memory & Language*, *41*, 416-426.
- RASTLE, K., & COLTHEART, M. (1999). Serial and strategic effects in reading aloud. *Journal of Experimental Psychology: Human Perception & Performance*, *25*, 482-503.
- RASTLE, K., & DAVIS, M. H. (2002). On the complexities of measuring naming. *Journal of Experimental Psychology: Human Perception & Performance*, *28*, 307-314.
- RASTLE, K., HARRINGTON, J., COLTHEART, M., & PALETHORPE, S. (2000). Reading aloud begins when the computation of phonology is

- complete. *Journal of Experimental Psychology: Human Perception & Performance*, **26**, 1178-1191.
- SCHNEIDER, W. (1988). Micro Experimental Laboratory: An integrated system for IBM PC compatibles. *Behavior Research Methods, Instruments, & Computers*, **20**, 206-217.
- TREIMAN, R., MULLENNIX, J., BIJELJAC-BABIC, R., & RICHMOND-WELTY, E. D. (1995). The special role of rimes in the description, use, and acquisition of English orthography. *Journal of Experimental Psychology: General*, **124**, 107-136.
- VAN SELST, M., & JOLICEUR, P. (1994). A solution to the effect of sample size on outlier elimination. *Quarterly Journal of Experimental Psychology*, **47A**, 631-650.
- ZORZI, M. (2000). Serial processing in reading aloud: No challenge for a parallel model. *Journal of Experimental Psychology: Human Perception & Performance*, **26**, 847-856.
- ZORZI, M., HOUGHTON, G., & BUTTERWORTH, B. (1998). Two routes or one in reading aloud? A connectionist dual-process model. *Journal of Experimental Psychology: Human Perception & Performance*, **24**, 1131-1161.

NOTES

1. It should be noted here that the LEX model currently has no semantics. Kwantes and Mewhort (1999) discuss LEX as a model of retrieval from semantic memory, and it is their terminology that is used here.
2. The purpose for restricting the study to monosyllabic words was that the PDP, CDP, and DRC models read only monosyllabic words.

3. Kučera–Francis frequencies were not significantly different for regular and irregular words in Position 2 [4.3 vs. 2.9, $t(66) = 1.2$, $p = .24$], nor were they significantly different for regular and irregular words in Position 3 [6.1 vs. 2.8, $t(34) = 1.2$, $p = .23$].

4. Items were not matched across position for length. There was a significant difference in length between Position 2 and Position 3 items [4.6 vs. 4.9, $t(102) = 2.23$, $SEM = .17$, $p < .05$] such that Position 3 items were longer than Position 2 items. In all item analyses, therefore, length was a covariate.

5. We are grateful to Marco Zorzi for carrying out these analyses.

6. The LEX model is different from the other three models tested in that responses for each item are different from one simulation to the next. Thus, 10 simulations were run with LEX and item response times are the mean response times from all simulations. Data for LEX were trimmed using a maximum criteria cutoff of 2,000, as recommended by P. Kwantes (personal communication, August 15, 2000).

7. The same test set submitted to each of the other models was also submitted to DRC. Removing items excluded in each of the other simulations did not affect the DRC simulation results; an analysis of all 104 items is thus reported.

8. There were more Position 2 pairs than Position 3, which could lead to more power to detect a regularity effect for Position 2 than for Position 3 irregulars. To address this concern, 18 Position 2 items were randomly sampled and analyzed with the 18 Position 3 items. This was done 10 times, and in all cases but two there was a significant interaction; numerically, in all cases the regularity effect was greater for Position 2 items than Position 3.

APPENDIX
Stimuli, Frequency, and Consistency Values

Item	Consistency	CELEX Frequency	Neighborhood Size	K-F Frequency
Position 2: Irregular Items				
bind	0.48	33	12	4
chaff	–	27	1	0
comb	–0.58	78	6	6
daft	–	39	5	0
dealt	0.17	71	1	22
dearth	0.45	16	1	3
douche	–	6	1	0
gauche	–	18	1	1
ghoul	–0.43	12	0	1
hearth	–0.6	78	4	4
hood	–0.5	74	11	7
knoll	0.19	31	1	2
leapt	–0.39	38	3	2
learnt	–	81	1	0
mauve	–0.56	80	0	1
monk	–0.05	67	5	16
mow	–	8	19	0
nook	–	14	7	0
pearl	–0.27	90	1	1
puss	–	12	12	0
rind	–	21	10	0
rook	–	1	11	0
shove	–0.5	33	4	2
sieve	–0.88	2	40	1
soot	–0.74	2	34	1
wad	–0.16	45	18	0
waltz	0.23	2	13	1
wan	–0.15	41	19	2
wand	–0.23	28	11	1
warp	0.33	38	10	4

APPENDIX (Continued)

Item	Consistency	CELEX Frequency	Neighborhood Size	K-F Frequency
wart	0.35	14	21	11
wasp	0.12	42	7	2
wharf	0.33	50	0	4
yearn	0.2	6	2	1
Position 2: Regular Items				
bluff	0.88	91	1	8
cheer	0.66	74	3	8
clutch	0.7	82	1	5
dale	0.76	35	16	5
doss	–	2	12	0
drake	0.9	31	3	2
drape	0.84	8	4	0
gleam	0.83	82	1	4
grouch	–	3	1	0
haunch	–	6	2	0
hiss	0.92	44	7	2
lull	0.68	38	10	2
lurch	0.47	30	2	3
mEEK	0.72	36	7	10
mop	0.85	44	16	3
muff	0.49	10	8	1
nick	0.88	27	12	25
notch	0.99	29	1	6
pod	0.84	40	18	3
pulp	0.52	70	4	5
rhyme	0.73	35	1	3
roost	0.75	19	2	1
shrine	0.9	78	3	7
snap	0.86	78	6	12
spruce	0.73	38	0	5
wade	0.79	13	15	2
wean	–	6	8	0
weave	0.86	39	2	4
weep	0.75	29	9	14
whiff	0.85	46	0	1
whirl	0.89	31	1	3
wick	0.92	30	12	4
winch	–	20	6	0
yam	–	8	12	0
Position 3: Irregular Items				
beige	–0.32	51	0	1
brook	–0.05	71	3	3
crook	–0.05	87	4	3
crow	–0.07	44	6	2
dreamt	–	14	2	0
grind	0.37	43	2	2
plaid	–0.69	49	2	1
scone	–	8	5	0
stead	–0.1	53	5	5
stealth	0.65	26	0	5
stow	–	7	6	0
swab	–	12	9	0
swamp	–0.46	78	3	5
swan	–0.27	84	8	3
swap	–0.42	17	9	2
swarm	0.42	38	2	3
swear	–0.53	64	4	10
tread	–0.19	45	6	5

APPENDIX (Continued)

Item	Consistency	CELEX Frequency	Neighborhood Size	K-F Frequency
Position 3: Regular Items				
bless	0.94	46	1	9
boost	0.75	70	2	15
claw	0.75	44	9	1
cleft	0.71	49	2	2
drudge	–	9	3	0
grunt	0.98	64	3	2
perch	0.75	38	3	1
scout	0.71	53	6	8
scrape	0.78	33	1	3
sling	0.9	44	8	1
slouch	0.88	9	1	1
slump	0.85	88	5	8
smug	0.63	43	4	7
smut	0.85	12	4	0
spleen	0.77	24	0	2
swoop	0.92	19	5	2
swung	0.45	75	3	48
teak	–	10	9	0

Note—Consistency values are listed only for items in CDP model corpus. K-F, Kučera–Francis.

(Manuscript received September 21, 2000;
revision accepted for publication April 26, 2002.)