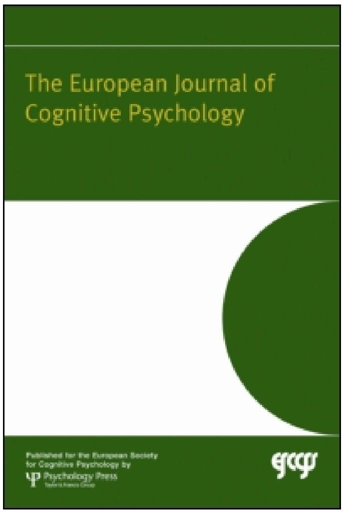

European Journal of Cognitive Psychology
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713734596

Rules versus statistics in reading aloud: New evidence on an old debate

Conrad Perry[a]; Johannes C. Ziegler[b]; Mario Braun[c]; Marco Zorzi[d]
[a] Swinburne University of Technology, Hawthorn, Victoria, Australia [b] Aix-Marseille Université, Aix-en-Provence, and Centre National de la Recherche Scientifique, Paris, France [c] Freie Universität Berlin, Berlin, Germany [d] Dipartimento di Psicologia Generale and Center for Cognitive Science, Università di Padova, Padova, Italy

First published on: 06 January 2010

To cite this Article Perry, Conrad , Ziegler, Johannes C. , Braun, Mario and Zorzi, Marco(2010) 'Rules versus statistics in reading aloud: New evidence on an old debate', European Journal of Cognitive Psychology, 22: 5, 798 — 812, First published on: 06 January 2010 (iFirst)

To link to this Article: DOI: 10.1080/09541440902978365

URL: http://dx.doi.org/10.1080/09541440902978365

EUROPEAN JOURNAL OF COGNITIVE PSYCHOLOGY
2010, 22 (5), 798–812

Psychology Press
Taylor & Francis Group

Rules versus statistics in reading aloud: New evidence on an old debate

Conrad Perry
Swinburne University of Technology, Hawthorn, Victoria, Australia

Johannes C. Ziegler
Aix-Marseille Université, Aix-en-Provence, and Centre National de la Recherche Scientifique, Paris, France

Mario Braun
Freie Universität Berlin, Berlin, Germany

Marco Zorzi
Dipartimento di Psicologia Generale and Center for Cognitive Science, Università di Padova, Padova, Italy

Nonword reading performance, that is, the ability to generate plausible pronunciations to novel items, has probably been the hardest test case for computational models of reading aloud. This is an area where rule-based models, such as the Dual-Route Cascaded (DRC) model, typically outperformed connectionist learning models. However, what is the evidence that people apply rules when reading nonwords? This was investigated in German. Nonwords were created that allowed us to test whether people apply an abstract rule to determine vowel length or whether they would be sensitive to the statistical distribution of vowel length in the mental lexicon. The human data showed a great amount of variability in nonword pronunciations. Simulations of these nonwords, where the DRC was contrasted with a fully implemented and freely available German version of the connectionist dual process model (German_CDP+), a model that learns the statistical mapping between spelling and sound, showed that CDP+ provided a better account of the

Correspondence should be addressed to Conrad Perry, Faculty of Life and Social Sciences (Psychology), Swinburne University of Technology, John Street, Hawthorn, 3122 VIC, Australia. E-mail: ConradPerry@gmail.com

This research was partly supported by grants from the Australian Research Council (DP0985815) to CP and the European Research Council (210922-GENMOD) to MZ. JZ was supported by an Alexander-von-Humboldt fellowship. We wish to thank Padraic Monaghan and Kathy Rastle for their insightful comments. The German_CDP+ is available for download at http://ccnl.psy.unipd.it/CDP.html

http://www.psypress.com/ecp DOI: 10.1080/09541440902978365

data than the DRC. These results support the view that rule based models may simply approximate patterns of language use rather than provide an accurate description of the underlying cognitive machinery.

Keywords: Computational modelling; Connectionism; German; Reading Aloud; Rules.

Whether human language processing (and, more generally, cognition) is best described by rule-based or connectionist models is one of the most fundamental and long standing debates in the cognitive sciences. The traditional (Chomskian) view that productivity in language is based on abstract symbolic rules (e.g., Clahsen, 1999; Marcus, Brinkmann, Clahsen, Wiese, & Pinker, 1995) has been challenged by statistical learning (a.k.a. connectionist) accounts in which "rule-governed" behaviour is assumed to simply be a description of patterns of language use that occur, rather than a description of the cognitive processes that cause it (e.g., Elman, 1993; Rumelhart & McClelland, 1986; Gupta & Touretzky, 1994). The issue, however, is far from being settled (cf. McClelland & Patterson, 2002; Pinker & Ullman, 2002), with much of the debate being focused on "quasiregular" domains (Plaut, McClelland, Seidenberg, & Patterson, 1996), whereby a limited number of idiosyncratic exemplars (i.e., exceptions) coexist with a large number of "regular" exemplars that can be described by a set of rules.

A well-known example of a quasiregular domain is the relationship that exists between spelling and sound at the sublexical level (i.e., the sublexical process used when reading aloud). The ability to learn the spelling-to-sound mappings and to generalise this knowledge to novel orthographic stimuli (i.e., nonwords) has been ascribed both to rule-based (Coltheart, Curtis, Atkins, & Haller, 1993; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001) and to connectionist learning systems (Plaut et al., 1996; Seidenberg & McClelland, 1989; Zorzi, Houghton, & Butterworth, 1998). The aim of the present study is to offer new insights into the rules versus connections debate by looking at a specific aspect of German orthography. In particular, in their implementation of the German DRC model, Ziegler, Perry and Coltheart (2000) noticed the existence of a simple abstract rule that can be used to predict, for a very large number of words, whether a vowel is likely to be pronounced long or short. This "metarule" takes into account higher level abstract information about letters—that is, whether a letter is a consonant or vowel and the number of consonants that occur after single letter vowels. If there is more than one consonant (i.e., VC+ words; e.g., *Milch* [milk]), the vowel is typically pronounced short. If there is only one consonant (i.e., VC words; e.g., *rot* [red]), the vowel is typically pronounced long. If such metarules exist, it is hard to see how connectionist models that only use simple associative mechanisms (e.g., Zorzi et al., 1998) could learn them.

The German DRC (Ziegler et al., 2000) directly implements the metarule in its set of rules. It therefore makes a strong prediction about German nonword reading: people should be sensitive to this metarule, and the sensitivity should be basically all or nothing. Thus, given a set of nonwords that orthogonally manipulates the number of consonants after the vowel, according to DRC, people should give pronunciations that use long vowels if there is only one consonant and short vowels if there are more than one.

In contrast, nonword reading with the connectionist dual process model (CDP+; Perry, Ziegler, & Zorzi, 2007) is based on a simple two-layer connectionist network that is sensitive to the frequency at which letters and sounds are associated in the lexicon (i.e., a *consistency* metric; e.g., Treiman, Mullennix, Bijeljac-Babic, & Richmond-Welty, 1995). This type of network has no way of learning complex metarules, and therefore predicts that the pattern of sensitivity should generally reflect the statistical distribution of simple letter–sound correspondences (see Appendix A).

A more specific prediction CDP+ makes is with respect to nonwords and how orthographically similar they are to real words. If sequences of graphemes commonly occur together, CDP+ may become sensitive to such co-occurrences over and above simple grapheme–phoneme frequencies, and this may affect the behaviour of the model. One such measure of word-likeness is the frequency at which a nonword shares its *body* (i.e., the orthographic equivalent of the rime) with other words. When conflict exists between the pronunciation that could be derived from the most common bodies and the pronunciation that could be derived from the most common graphemes, CDP+ is more likely to use a pronunciation based on the body when the body is frequent (see Perry et al., 2007; Zorzi et al., 1998). This makes the interesting prediction that CDP+ and possibly people might give different responses as a function of whether the body of a nonword is extant or not.

To examine these predictions, we chose a set of nonwords that orthogonally manipulated two factors. The first was simply the number of consonants after the vowel, which was either one or more than one. The second was whether the nonword used an extant or nonextant body. If people perfectly complied with the rules when reading nonwords aloud, they should give short vowel pronunciations to all VC+ nonwords regardless of whether they have extant or non-extant bodies. We also carried out simulations with the German DRC (Ziegler et al., 2000) and a newly developed German version of CDP+ to investigate which of the models would more closely capture the human data.

METHODS

Participants

Thirty-four undergraduate students from the University of Eichstätt, Germany, participated in the study.

Stimuli

Eighty nonwords were used. They were divided into four groups based on whether they had an extant body and whether they had one or more than one consonant after the vowel. All had single letter vowels. The nonwords were chosen such that the onsets were matched quadruplet-wise across the four groups, as were the vowels. The average letter length of the nonwords was matched within VC (extant: 3.50; nonextant: 3.50) and VC+ (extant: 4.95; nonextant 4.90) groups. Forty further nonwords that did not have a VC or VC+ body were added to the stimuli set and used as fillers.

Procedure

The stimuli were presented in the centre of the screen and participants were asked to read them aloud as quickly and as accurately as possible. The stimuli disappeared from the screen as soon as the voice-key registered that the participant had made a response. The third author marked the stimuli as having a long or short vowel. The stimuli were presented in a 24 point Courier font in black on a white background.

RESULTS

Human data

All fillers were discarded from the analysis. One item (*Witch*) was removed from the analysis because most participants gave the English loan-word pronunciation. A further 6.03% of the remaining responses were removed from the reaction times (RTs) and considered errors because participants gave an implausible nonword pronunciation for them or stuttered whilst producing them. The plausibility of nonword pronunciations was judged by the data-coder. Note that in a transparent writing system, such as German, the decision about what counts as a plausible pronunciation is obvious because there is very little ambiguity about how letters should be pronounced. Thus, most errors were due to lexicalisations or visual confusions. For the RT analysis, items with response times greater than

1200 ms or less than 100 ms were removed from the analysis (0.85%) as were any remaining responses that were more than 3 *SD*s away from the mean RTs calculated for each participant (0.49%).

The results showed that participants gave more short vowels with the VC+ versus VC nonwords (VC: 49.46%; VC+: 94.63%), $F1(1, 33) = 309.76$, $p < .001$; $F2(1, 18) = 131.50$, $p < .001$. Participants also appeared to give more short vowels in the nonextant body condition, $F1(1, 33) = 27.52$, $p < .001$; $F2(1, 18) = 6.47$, $p < .005$, although this appeared to be restricted mainly to the VC nonwords (VC+, extant vs. nonextant: 94.73% vs. 94.52%; VC, extant vs. nonextant: 40.54% vs. 58.58%), causing a significant interaction, $F1(1, 33) = 39.19$, $p < .001$; $F2(1, 18) = 5.53$, $p < .05$. The results appear in Table 1.

In terms of RTs, participants read aloud the VC+ nonwords more slowly than the VC nonwords (VC+: 627.37 ms; VC: 589.86 ms), $F1(1, 33) = 53.33$, $p < .001$; $F2(1, 18) = 31.53$, $p < .001$. There was also a main effect of extant body status that was significant by participants but not items, $F1(1, 33) = 17.92$, $p < .001$; $F2(1, 18) = 2.42$, $p = .14$ (nonextant: 613.66 ms; extant: 602.61 ms). The interaction, where VC nonwords showed a smaller effect than VC+ nonwords (5.51 ms vs. 19.46 ms), was significant only by participants, $F(1, 33) = 4.69$, $p < .05$; $F2 < 1$. The results appear in Table 2.

In terms of errors, most occurred in the VC+ group with nonextant bodies (VC/extant body: 2.50%; VC+/extant body: 3.68%; VC/nonextant body: 3.24%; VC+/nonextant body: 11.18%). Accordingly, the main effects of number of consonants, $F1(1, 33) = 12.69$, $p < .005$; $F2(1, 18) = 13.56$, $p < .005$, and extant body status, $F1(1, 33) = 23.04$, $p < .001$; $F2(1, 18) = 9.35$, $p < .01$, were significant, as well as their interaction, $F1(1, 33) = 13.94$, $p < .005$; $F2(1, 18) = 9.80$, $p < .01$.

TABLE 1
Mean percentage of short vowels given on the nonwords for the models and humans, as a function of extant body status and body type

	Participant(s)				
	Human		*CDP+*	*CDP+ (without umlauts)*	*DRC*
Body type	*Mean*	*SD*	*Mean*	*Mean*	*Mean*
Nonextant					
VC	59	24	45	56	0
VC+	95	5	80	88	100
Extant					
VC	41	25	30	38	0
VC+	95	8	75	88	100

TABLE 2
Mean reaction times for the models (cycles) and humans (ms), as a function of extant body status and body type

	Participant(s)					
	Human		*CDP+*		*DRC*	
Body type	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
Nonextant						
VC	593	119	129	18	136	7
VC+	638	146	153	20	188	24
Extant						
VC	587	115	129	18	131	13
VC+	618	131	149	21	181	23

To summarise these results, participants were sensitive to the number of consonants after the vowel, but this sensitivity was not all or nothing, with long vowels only being used around half the time in the VC condition. Thus, it appears that our participants did not consistently use a consonant counting rule to determine vowel length. In addition, it also appears that our participants were more likely to give long vowel pronunciations with VC nonwords if they had extant bodies. Thus, the frequency at which a body occurred affected responses.

Simulation study

We conducted simulations with a German version of the rule-based DRC and a German version of the connectionist learning model CDP+ (*German_CDP+*) in order to investigate which of the models would capture more closely the human data.

German_CDP+. A German version of CDP+ was constructed (the executable version can be downloaded at http://ccnl.psy.unipd.it/CDP.html). We used the lemma database from CELEX (Baayen, Piepenbrock, & van Rijn, 1993) that was also used in the German DRC (Ziegler et al., 2000). Identical choices as in DRC were also made when there were potential options (i.e., -ss was used and not β, and capital and lower case letters were treated identically). The model was trained as in Perry et al. (2007) except for the following differences: (1) The pretraining phase used the set of German spelling-to-sound correspondences (based on phonics teaching programmes) from Hutzler, Ziegler, Perry, Wimmer, and Zorzi (2004), modified slightly to fit into the grapheme-based coding scheme. (2) We trained the model for 500 cycles rather than 100 because the database is around one-fifth the size of the English one. (3) We allowed an attentional window of four letters instead of three because the largest German grapheme has four letters (i.e., -tsch). The graphemes used in the graphemic buffer of the model appear in Appendix B. Although it is likely that parameters used may differ in different languages (see, e.g., Perry & Ziegler, 2002; Ziegler, Perry, & Coltheart, 2003), there is currently not enough data to effectively optimise the parameters of CDP+_German. We therefore used exactly the same set of parameters as Perry et al. (2007). To make sure that the results of the model were not due to idiosyncratic factors with respect to the order of presentation of the stimuli in training, we repeated the procedure 10 times and chose a median model in terms of the proportion of short/long vowels given.

German DRC. The German DRC is fully described in Ziegler et al. (2000). The parameters used are identical to the English version reported in Coltheart et al. (2001).

Simulation results

All errors were removed from the analysis (DRC: 3; German_CDP+ : 0). As can be seen from Table 1, the two different models make quite different predictions as to the expected number of long and short vowels that should be given in the different categories. Not surprisingly, DRC predicted that the metarule should be used—in fact, 100% of the time. Alternatively, German_CDP+ predicted that we should find many VC nonwords for which people give a short vowel answer (37.5%). German_CDP+ also produced more short vowels with VC nonwords in the nonextant versus extant condition (nonextant: 45%; extant 30%). However, the interaction between vowel length and extant body status was not significant, $F < 1$. An uncorrected two-tailed paired samples t-test examining extant body status with just the VC nonwords was marginally significant, $t(19) = 1.83$, $p = .083$.

We also examined generalisation performance of the model without umlauts (i.e., we excluded 16 of the 80 nonwords that contained the vowels ä, ö, and ü). This was done because the distribution of umlauts and the vowels they map onto is very different in monosyllabic as opposed to multisyllabic words. For example, the umlauted letter *ä* most commonly maps onto short vowels in the first syllable of disyllables, but maps onto short vowels in monosyllables (disyllables: ä: 57.1%; ö: 42.7%; ü: 62.9%; monosyllables: ä: 25%; ö: 36.4%; ü: 60%). Given that we used a monosyllabic database for training, the model may therefore be affected by potentially incorrect statistical information presented to it. As can be seen from Table 1, when nonwords with umlauts are removed from the testing set, the proportion of short vowel responses of German_CDP+ increased in all four categories.

In summary, when the human data is compared with German_CDP+ and DRC, the distribution of short and long vowels given by participants was much more similar to German_CDP+ than DRC, especially with respect to VC nonwords. That was confirmed by calculating root mean squared error differences based on the mean proportion of times participants gave long or short vowels in the four different categories and the mean results given by the models (including the umlauts) (German_CDP+: .15; DRC: .36; note that smaller numbers indicate a better fit). The actual difference between the percentages of long/short vowels given with VC nonwords in the non-extant and extant body groups was almost the same in the human (18%) and German_CDP+ (15%) data.

In terms of RTs, German_CDP+ predicted that there should be no significant effect of extant body status, $F < 1$ (nonextant: 141.1 cycles; extant: 139.0), whereas DRC did predict a significant effect: $F(1, 16) = 6.18$, $p < .05$ (nonextant: 162.1 cycles; extant: 155.6 cycles). Both models predicted a difference between the VC and VC+ nonwords—that is, that there will be a length effect in nonword reading, German_CDP+: $F(1, 19) = 31.23$, $p < .001$; DRC: $F(1, 16) = 124.58$, $p < .001$. Neither model predicted an interaction, German_CDP+: $F < 1$; DRC: $F < 1$.

DISCUSSION

The results of the study are clear. Nonword pronunciations of German speakers appear to be much closer to what might be expected based on the statistical distribution of letter–sound relationships in a lexical corpus (see Appendix A) than the use of a metarule that relies on consonant counting. The human data are generally consistent with German_CDP+ because humans showed the same kind of variability that was predicted by the model. In fact, German_CDP+ uses only a simple two-layer network that cannot learn complex nonlinear rule-like patterns. Thus, even if metarules might be useful at a descriptive level, people seem to be more sensitive to the subtle statistical patterns present in the lexicon rather than all-or-none rules, and this is what German_CDP+ predicts.

In terms of RTs, the main result of interest was that whilst there was an effect of extant body status, it was only significant by participants. This result is essentially the same as that found by Ziegler, Perry, Jacobs, and Braun (2001). The present study extends this finding by showing a significant difference in nonword generalisation performance (as measured by number of short/long vowels given) and error rates as well. Thus, extant body status does strongly affect processing, but RTs are not especially sensitive to it.

Apart from the experiment reported here, there are a small number of other data sets that can be used to examine the model. One such set (Ziegler et al., 2001) examined the cross-language length effect, where German readers show a larger length effect than English readers, both on words and nonwords. In simulations of this effect, Perry and Ziegler (2002) showed that, at least with nonwords, the DRC produced such an effect due to the serial nature of the assembled phonology route. Perry and Ziegler demonstrated that the reason for this was because, when assembling VC nonwords, the DRC initially produces a long vowel when the VC part becomes available but that this later gets revised to a short vowel when the VCC part becomes available. This revision process slows processing down in German with VCC nonwords in two ways. First, it means that the correct vowel becomes available relatively late in the assembly processes. Second, when the correct vowel becomes available, there is spurious activation left from a previously incorrect vowel being assembled. Due to these two factors, the German DRC is slower at producing longer nonwords than shorter ones compared to the English DRC where this particular pattern does not exist.

We tested whether German_CDP+ would produce the German-English pattern in the same way as the DRC. The results showed that whilst German_CDP+ displays a similar pattern of revision due to the left–right parsing of graphemes, the revision process is not strong enough to cause a larger length effect in German than English. Whilst it may be possible to reparameterise the model to accentuate this effect, we did not do this because there is a more parsimonious explanation. In particular, in Perry et al. (2007), we suggested that the parsing of graphemes into the graphemic buffer requires focused spatial attention. In this respect, an important difference between German and English is that the grapheme parsing mechanism requires a three-letter attentional window in English but a four-letter window in German. If we assume that using a four-letter window is more attention demanding than using a three-letter window, this would affect the speed at which graphemes can be parsed. In particular, it would cause graphemes to become available more slowly, thus creating a greater length effect in German compared to English. Whether or not this hypothesis is correct is an empirical question that can be tested. If the hypothesis is correct, the length effect in languages that do not require four-letter attentional windows should be weaker than that found in German.

Apart from German_CDP+, there may be other types of model that may be able to explain the data if modified. In terms of the DRC, one change that might be used would be to get rid of the metarule that we proposed in Ziegler et al. (2000) and use a set of all-or-nothing grapheme-phoneme conversion (GPC) rules based on extant grapheme–phoneme correspondences instead.

However, this would lead to an explosion of fairly complex context-sensitive rules. This is because each vowel would need to have a large set of context-sensitive rules that determine its pronunciation (basically, all existing combinations of consonants that follow the vowel). For example, with just single letter vowels with either one or more than one consonant after them such as those examined here, there would need to be 71 VC and 303 VCC rules, as this is how many unique extant sequences of VC and VCC letters there are. Furthermore, if graphemes and not just the two letters after the vowel are important (as the results from our extant body manipulation suggest), but only letters are used to determine context (as is the case with the DRC), then the rules needed would have to include some with five context letters. Without these, some bodies that use five letters in their codas but have only two graphemes (e.g., -*atscht*, which exists in words like *klatscht* [he/she claps]) could not be distinguished from bodies that use four letters but only one grapheme. Such a large set of context-sensitive rules would be needed since the pattern of results whereby single-letter vowels often receive both long and short vowel responses cannot be explained by the frequency of context-free grapheme–phoneme correspondences—it can only be explained when consonantal context is taken into account.

Although using a large number of rules to predict nonword pronunciations may not be especially parsimonious, based on the database, 92% of words with VC+ bodies and 25% of words with VC bodies used a short vowel, and, thus, a set of rules based on extant words that use a large amount of consonantal context may come close to approximating the overall difference between VC and VC+ nonwords with extant bodies. Despite this, a problem with using large numbers of rules and not a metarule is that it is not clear how nonwords with nonextant bodies could be read aloud. These nonwords also showed a rather mixed pattern in terms of the proportion of short and long vowels given, with VC+ nonwords almost always being assigned short vowels, but VC nonwords being assigned short vowels only around half the time. This has implications for both the VC and VC+ nonwords.

In terms of VC+ nonwords with nonextant bodies, these may create problems because although their bodies were nonextant, they generally had extant VC sequences within them. For example, whilst the -enst in *spenst* does not exist as a body, the -en does (e.g., *Gen* [gene]). Thus, if a set of GPC rules with consonantal context sensitivity was used, the -en would be assembled with a VC rule, which would predict that nonwords like *spenst* should have a long/short vowel distribution more similar to that of VC words than VC+ words, which is not what was found. One way around this would be to mark all of the VC rules as "end" rules (i.e., rules that only apply at the end of words). However, if this was done, the rules would simply amount to a list of extant bodies, and, in addition, it would still not be possible to distinguish between vowels in VC and VC+ nonwords with nonextant bodies.

In terms of VC nonwords with nonextant bodies, these may create problems because the relationship between the final consonant and the vowel does not exist and hence could not be learnt. Thus there is simply no way to predict a distribution with VC nonwords with nonextant bodies. In this case, since single letter vowel grapheme–phoneme correspondences map onto short vowels much more often than long vowels, VC nonwords with nonextant bodies should almost always be given short vowels, which is not what was found.

Given that all-or-nothing contextually sensitive rules could not predict the data, a further possibility would be to allow probabilistic weighting of rules. Thus, given the presentation of the same grapheme, different grapheme–phoneme rules could be selected in different circumstances based on lexical statistics. Using probabilistic rules might seem intuitively reasonable and could perhaps explain some of the present data, but it is not without its own problems. One in particular is that, at least for some types of nonwords in English, people almost always choose regular pronunciations (see Andrews & Scarratt, 1998), and thus a probabilistic rule system might underestimate the number of regular pronunciations with some types of nonword.

Even if probabilistic rules were added and they did not cause too many nonregular pronunciations, it is not clear that simple grapheme–phoneme correspondences would be enough to capture effects that occur at the rime-body level, such as consistency effects with words (e.g., Jared, 2002). Thus, one might need to add rules of different grain sizes, in addition to suggesting some method by which small (e.g., phoneme–grapheme) and large (e.g., body–rime) rules would be selected when in competition with each other. This would be a very big change to the present model.

Aside from practicalities of using a larger number of probabilistic context-sensitive rules, radically changing the rules of DRC would clearly soften some of the main tenets of a rule-based approach and some of the main advantages, such as its rather simple and straightforward solution to explaining how people read words and generalise to nonwords. Thus, theoretically, updating the DRC to handle probabilistic data via the use of probabilistic rules would mean that one of the great advantages of a rule-based approach would be lost.

An interesting alternative to the models we have discussed so far would be a lexical analogy plus rules model (Andrews & Scarratt, 1998; see Campbell, 1983, for such a model of spelling; and Albright & Hayes, 2005, for an implemented computational model in a slightly different domain). This type of a model first attempts to read nonwords by analogy to existing words. If the analogical process fails, a set of rules are used. This means that the less word-like a nonword is (and hence the harder it is to construct a pronunciation via analogy), the more likely rules are to be used when reading it (see Andrews & Scarratt, 1998, for a discussion). In our case, this means that VC nonwords with nonextant bodies should be read aloud with long vowels more often than VC nonwords with extant bodies, because the model would be more likely to default to a rule that uses a long vowel in the nonextant case (assuming a metarule is used). This is the opposite of what CDP+ predicted and the opposite of what was found in the human data. Again, this suggests that CDP+ does a better job at predicting the qualitative data pattern than other types of model.

A second type of analogy model[1] worth considering is that of van den Bosch and Daelmans (1998). This type of model is computationally implemented and works by breaking words down into all the possible fixed length letter sequences they have (seven-letter sequences are used in the current model). The sequences themselves consist of a central letter and context letters to the left and right of it and are stored with the phoneme that cooccurs with the central letter. The pronunciation of nonwords is then constructed by matching the sequences which can be created from the nonword with the most commonly occurring sequences that have been stored based on a nearest-neighbour algorithm. In its present form, it is unlikely that this type of analogy model could capture the general pattern found here, since the number of context letters to the right of the vowel is important in determining the vowel pronunciation, and the current three-letter context is not sufficient to capture the distinction between VC and VC+ nonwords that use a coda grapheme after the vowel that is three or four letters long. This could potentially be fixed by allowing a five-letter context, but this would have the undesirable effect of increasing the already high error rate of the model. Given the lack of rigorous testing of this model on even the most basic effects (e.g., length), further discussion of these possibilities seems premature.

In conclusion, our results support the view that, in some instances, rule-based models may simply approximate patterns of language use rather than necessarily provide a description of how cognitive processing occurs (e.g., Gupta & Touretzky, 1994). The results here show that phonological assembly in German_CDP+, which uses a network that learns statistical patterns between spelling and sound, provides a better description of human knowledge of productive spelling–sound relationships than rule-based models, even in a case that could easily be handled by a single abstract and highly predictive rule. Finally, the results demonstrate that detailed analyses of nonword pronunciations beyond simple accuracy measures (e.g.,

[1] We thank Padraic Monaghan for this suggestion.

Andrews & Scarratt, 1998) can provide very constraining datasets even in the case of a transparent orthography like German and support our programme to extend the CDP modelling approach to German and other European languages. Finally, we hope that the free availability of German_CDP+ will stimulate new research on reading aloud in German, in both within- and cross-language studies.

REFERENCES

Albright, A., & Hayes, B. (2005). Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition*, *90*, 119–161.

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

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). *The CELEX Lexical Database* [CD-ROM]. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Campbell, R. (1983). Writing nonwords to dictation. *Brain and Language*, *19*, 153–178.

Clahsen, H. (1999). Lexical entries and rules of language: A multidisciplinary study of German inflection. *Behavioral and Brain Sciences*, *22*, 991–1060.

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

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

Elman, J. (1993). Learning and development in neural networks: The importance of starting small. *Cognition*, *48*, 71–99.

Gupta, P., & Touretzky, D. S. (1994). Connectionist models and linguistic theory: Investigations of stress systems in language. *Cognitive Science*, *18*, 1–50.

Hutzler, F., Ziegler, J. C., Perry, C., Wimmer, H., & Zorzi, M. (2004). Do current connectionist learning models account for reading development in different languages? *Cognition*, *91*, 273–296.

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

Marcus, G. F., Brinkmann, U., Clahsen, H., Wiese, R., & Pinker, S. (1995). German inflection: The exception that proves the rule. *Cognitive Psychology*, *29*, 189–256.

McClelland, J. L., & Patterson, K. (2002). Rules or connections in past-tense inflections: What does the evidence rule out? *Trends in Cognitive Sciences*, *6*, 464–465.

Perry, C., & Ziegler, J. C. (2002). Cross-language computational investigation of the length effect in reading aloud. *Journal of Experimental Psychology: Human Performance and Perception*, *28*, 990–1001.

Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: The CDP+ model of reading aloud. *Psychological Review*, *114*, 273–315.

Pinker, S., & Ullman, M. T. (2002). The past and future of the past tense. *Trends in Cognitive Sciences*, *6*, 456–463.

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.

Rumelhart, D., & McClelland, J. L. (1986). On learning the past tenses of English verbs. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Vol. 2. Psychological and biological models* (pp. 216–271). Cambridge, MA: MIT Press.

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

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 den Bosch, A., & Daelmans, W. (1998). Do not forget: Full memory in memory-based learning of word pronunciation. In D. M. W. Powers (Ed.), *Proceedings of the Joint International Conference on New Methods in Language Processing and Computational Natural Language Learning* (NeMLaP3/CoNLL98), Sydney, Australia, January 1998, pp. 195–204.

Ziegler, J. C., Perry, C., & Coltheart, M. (2000). The DRC model of visual word recognition and reading aloud: An extension to German. *European Journal of Cognitive Psychology*, *12*, 413–430.

Ziegler, J. C., Perry, C., & Coltheart, M. (2003). Speed of lexical and nonlexical processing in French: The case of the regularity effect. *Psychonomic Bulletin and Review*, *10*, 947–953.

Ziegler, J. C., Perry, C., Jacobs, A., & Braun, M. (2001). Identical words are read differently in different languages. *Psychological Science*, *12*, 379–384.

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 and Performance*, *24*, 1131–1161.

APPENDIX A

Vowel pronunciation counts of all monosyllabic German words with a single letter vowel grapheme in the German CELEX lemma database (Baayen et al., 1993)

	Vowel pronunciation		
	Short	*Long*	*Total*
VC words			
Vowel			
A	8	47	55
Ä	0	7	7
E	9	11	20
I	17	5	21
O	9	40	49
Ö	0	7	7
U	9	36	45
Ü	0	7	7
Total	52	160	211
VC+ words			
Vowel			
A	251	19	270
Ä	3	2	5
E	116	9	125
I	135	1	136
O	113	24	137
Ö	4	4	8
U	106	3	109
Ü	15	3	18
Total	743	65	808

APPENDIX B

Graphemes used in the graphemic buffer of CDP+

Grapheme type	*Graphemes*
Onset	sch, kn, pf, ch
Vowel	auh, ieh, äu, oa, ie, ai, au, ei, eu, ee, aa, oo, ou, ah, eh, ih, oh, uh, äh, öh, üh
Coda	tsch, sch, tch, ch, ck, ff, ll, mm, ng, nn, pf, pp, rr, ss, th, ts, tt, tz, zz, dt, ph