Quantifying the reliance on different sublexical correspondences in German and English

Xenia Schmalz, Eva Marinus, Serje Robidoux, Sallyanne Palethorpe, Anne Castles, and Max Coltheart

ARC Centre of Excellence in Cognition and Its Disorders, Department of Cognitive Science, Macquarie University, Sydney, NSW, Australia

The type of sublexical correspondences employed during non-word reading has been a matter of considerable debate in the past decades of reading research. Non-words may be read either via small units (graphemes) or large units (orthographic bodies). In addition, grapheme-to-phoneme correspondences may involve context-sensitive correspondences, such as pronouncing an “a” as /ɔ/ when preceded by a “w”. Here, we use an optimisation procedure to explore the reliance on these three types of correspondences in non-word reading. In Experiment 1, we use vowel length in German to show that all three sublexical correspondences are necessary and sufficient to predict the participants’ responses. We then quantify the degree to which each correspondence is used. In Experiment 2, we present a similar analysis in English, which is a more complex orthographic system.

Keywords: Optimisation; Reading; Sublexical processing.

How print is converted to speech is an important question, both from a theoretical and a practical perspective. Sublexical translation processes have a central role in all current models of reading aloud (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Perry, Ziegler, & Zorzi, 2010; Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg & McClelland, 1989). The exact nature of this sound-to-speech conversion procedure, however, has been under considerable debate since the 1970s. In particular, the debate revolves around the question of whether this conversion relies predominantly on small units, such as graphemes, or larger units, such as orthographic bodies (e.g., “-ord”) (Andrews, 1982; Coltheart et al., 2001; Cortese & Simpson, 2000; Glushko, 1979; Jared, 2002). To a lesser extent, the literature has also drawn a distinction between context-sensitive and context-insensitive grapheme-to-phoneme correspondences (GPCs) and addressed the possibility that rather than relying purely on single-grapheme correspondences, in some cases the preceding or

---

Correspondence should be addressed to Xenia Schmalz, ARC Centre of Excellence in Cognition and Its Disorders, Cognitive Science, Macquarie University, Sydney, NSW, 2109, Australia. E-mail: xenia.schmalz@mq.edu.au

We are grateful to Angela Heine and Tila Brink for collecting the Berlin data for Experiment 1B. We also thank Petra Schienmann and Reinhold Klégl for their help with organising data collection at Potsdam University for Experiment 1B. We thank Johannes Ziegler for providing a list of German consistent words. Further thanks are due to Stephen Lupker, James Adelman and three anonymous reviewers for their helpful comments on earlier versions of this paper.

This article was written as part of XS’s doctoral dissertation under supervision of EM, AC and MC. SR conducted the data analyses and contributed to the write-up and revision of the manuscript, and SP scored the English data.

© 2014 Taylor & Francis
succeeding letters may provide a cue to the reader about the correct pronunciation of a grapheme (Perry, Ziegler, Braun, & Zorzi, 2010; Treiman, Kessler, & Bick, 2003; Treiman, Kessler, Zevin, Bick, & Davis, 2006).

Thus, the literature reports three different types of correspondences that may be involved in sublexical decoding: context-insensitive GPCs, context-sensitive GPCs and body–rime correspondences (BRCs). Here, we propose a mathematical model based on an optimisation procedure that will allow us to fit the degree of reliance on each of the three types of correspondences. We begin with two experiments in German, where the language structure allows us to assess the independent contribution of each of the three types of correspondences. In two further experiments, we apply the same methodology to the English grapheme “a”, which allows us to disentangle the reliance on context-sensitive GPCs compared to context-insensitive GPCs.

GPCs describe the relationship between graphemes and phonemes. The phoneme is the basic unit in spoken language, and a grapheme is the letter or letter cluster that corresponds to a single phoneme. The definitions of GPCs are straightforward in some cases; for example, the grapheme “b” always maps onto the phoneme /b/. This is an example of a context-insensitive GPC: regardless of the letters that precede or succeed the grapheme, its assigned phoneme does not change. However, this gets more complicated when we consider the GPC for a grapheme such as “a”. In English, context-insensitive correspondences would dictate that “a” should be pronounced as in “cat”. Using this correspondence, words like “was” and “false” would be considered irregular, meaning that the correct pronunciation is inconsistent with the GPC. Yet, upon closer inspection, the pronunciations of “was” and “false” are entirely predictable when the context of the grapheme “a” is taken into account: in “was”, the “a” is preceded by a “w”, which in most cases changes the pronunciation to /æ/, as in “wad” and “swan”. This context-sensitive correspondence can be written as “[w]a” → /æ/. The pronunciation of the vowel in “false” can be similarly predicted by a complex context-sensitive GPC, namely “[C]a[I][C]” → /əː/.

It is worth noting that these context-sensitive correspondences (CSCs) are still GPCs, as they relate a single grapheme (in this case, “a”) to the pronunciation of a single phoneme. Thus, GPCs can be subdivided into context-sensitive GPCs (“[w]a” → /æ/) and context-insensitive GPCs (“a” → /əː/).

The concept of GPCs is important for the classical computational model of the dual-route framework, the DRC (Coltheart et al., 2001). This model has a sublexical route which converts print to speech via a set of GPCs that are explicitly specified. The sublexical route contains some CSCs (N = 28—although the exact numbers vary according to the version of the DRC), but operates mostly on single-letter (e.g., “b” → /b/; N = 40) and multi-letter (e.g., “th” → /θ/; N = 165) context-insensitive GPCs.

There is also experimental evidence that stresses the importance of CSCs. One study reported the case of a patient with acquired surface dyslexia (Patterson & Behrmann, 1997): since this patient could not correctly read irregular words like “colonel” and “yacht”, it was thought that her lexical system was heavily damaged. However, not all irregular words were a problem: she was unimpaired with words that could be resolved by the context-sensitive “[w]a” → /æ/ correspondence, such as “wad” or “swan”. This demonstrates the presence of such a context-sensitive correspondence in the sublexical system. Furthermore, studies of non-word reading have shown that there is psychological reality to CSCs (Treiman et al., 2003, 2006): both adults and children tend to pronounce non-words such as “twamp” with the vowel as in “swan”, whereas control items such as “glamp” are pronounced via the context-insensitive GPC, → /æ/. This further suggests that the context-insensitive correspondence “a” → /æ/ does not fully reflect the strategies used during non-word reading.

In addition to context-insensitive and context-sensitive GPCs, readers have been shown to rely on BRCs. BRCs are the sublexical links between bodies and rimes, where bodies are defined as the vowel and optional final consonant(s) of a monosyllabic word (e.g., “ark” in the word “bark”). The rime is the phonological equivalent to the orthographic body. A linguistic analysis has shown that bodies are a reliable predictor of vowel
pronunciation in English (Treiman, Mullennix, Bijeljac-Babic, & Richmond-Welty, 1995).

Full reviews of the psychological reality of BRCs can be found elsewhere (Goswami & Bryant, 1990; Ziegler & Goswami, 2005). Most relevant in the current context are non-word reading studies addressing this issue because these allow for a systematic exploration of the non-lexical correspondences that participants rely on when lexical information is not available. In English, non-words can be created that would yield different responses depending on whether GPCs or BRCs are used. This is done by manipulating the regularity and consistency of the base word. A base word that conforms to the context-insensitive GPCs is said to be regular, whereas words that violate the correspondences are considered to be irregular. The concept of regularity only matters if reading occurs at least in part via GPCs. If non-lexical reading occurs only via BRCs, then the reliability (or lack thereof) of the GPC information should not influence reading at all; rather, only inconsistency of BRCs should affect reading latencies and accuracy (e.g., the two ways of pronouncing “ave” in “have” and “save”, see Ziegler, Stone, and Jacobs 1997).

Non-word reading studies that aim to estimate the reliance on GPCs versus BRCs can use the regularity and consistency of a base word to generate non-words that predict different responses depending on the types of correspondences that are used by the participant. Such studies are important, because non-word reading data can shed light on processing underlying sublexical information, while minimising confounds from lexical processing. Understanding this process has strong theoretical implications because sublexical print-to-speech conversion mechanisms play an important role in all prominent models of reading.

In order to disentangle the different sublexical processes that take place during reading, the first step is to create non-words for which different types of correspondences make different predictions. For example, from a regular and consistent word such as “fact”, the onset can be changed to create a non-word, for example, “ract”. In this case, both large and small correspondences make the same predictions for pronouncing this non-word. However, if we take an irregular, but consistent word, such as “talk”, and change the onset to create the non-word “talk”, we can use the readers’ pronunciations of this non-word to determine whether they relied on context-insensitive GPCs (in which case the item would be pronounced to rhyme with “talc”) or BRCs (where it would rhyme with “talk”). Such studies have shown that GPCs cannot fully account for the types of pronunciations that participants give to such non-words, but neither do BRCs (Andrews & Scarratt, 1998; Brown & Deavers, 1999; Perry, Ziegler, Braun, & Zorzi, 2010; Pritchard, Coltheart, Palethorpe, & Castles, 2012).

Thus, there is evidence for reliance on the three different types of print-to-speech correspondences, but there are still questions that remain to be answered. First, previous studies do not distinguish between the reliance on context-sensitive GPCs and BRCs. For example, if a participant pronounces the non-word “false” to rhyme with “false”, it may be that a context-sensitive correspondence, “[a][l]” → /o/ has been used to derive the pronunciation, rather than the BRC that “-alse” → /o:ls/. As will be discussed later, this is a problem in the English language, as BRCs and CSCs are confounded.

Second, even though such studies can establish the psychological reality of certain types of correspondences, examining between-item differences does not allow estimation of the relative degree to which each type of correspondence plays a role. As previous research has demonstrated the psychological reality of context-insensitive GPCs, context-sensitive GPCs and BRCs, it is likely that all three correspondence types help the sublexical route to determine the pronunciation of a non-word. How such a conflict between different types of correspondences may be resolved by the cognitive system is addressed in detail in the General Discussion. The possibility of parallel activation of several sublexical correspondences raises the question of whether it is possible to quantify the degree to which each plays a role in determining the pronunciation of a novel word, which is a natural next step after demonstrating a sublexical correspondence’s psychological reality. As discussed later, more sophisticated analyses are needed to estimate the relative importance of each type of correspondence.

In addition to establishing the psychological reality of different types of sublexical correspondences, a considerable body of research has explored cross-linguistic differences in the reliance on GPCs versus BRCs (Goswami, Gombert, & De Barrera, 1998; Goswami, Porpodas, & Wheelwright, 1997; Goswami, Ziegler, Dalton, & Schneider, 2003; Ziegler, Perry, Jacobs, & Braun, 2001; Ziegler, Perry, Ma-Wyatt, Ladner, & Schulte-
Körne, 2003). The psycholinguistic grain-size theory, a cross-linguistic theory of reading development and skilled reading, proposes that the degree of reliance on sublexical correspondences of different types varies across languages (Ziegler & Goswami, 2005). In particular, the reliance on BRCs has been reported to be stronger in English than German (Ziegler et al., 2001, 2003). This is argued to be because in English, large units (i.e., bodies) are a better predictor of the pronunciation of a word than GPCs (Treiman et al., 1995): for a word like “calm”, the pronunciation is inconsistent with the GPCs (“kælm”) but can be derived from its body neighbours (palm, balm, etc.). In German, on the other hand, the GPCs are highly reliable, meaning that there are few exceptions to the correspondences (Ziegler, Perry, & Coltheart, 2000); therefore, smaller units are the preferred grain size of German readers. In other words, there is a theoretical framework which predicts differences in the reliance on the units across languages. Therefore, it is desirable to develop a mathematical model quantifying the degree of reliance in different languages.

In summary, previous literature has shown reliance on three different types of correspondences in English: Context-insensitive GPCs, context-sensitive GPCs and BRC. The psycholinguistic grain-size theory proposes that the reliance on the different types of correspondences differs across languages (Ziegler & Goswami, 2005). In the present experiments, we introduce a new method of quantifying the reliance on each type of correspondence. In the first two experiments (1A and 1B), we used German non-words to assess the degree of reliance on each type of correspondence. In Experiments 2A and 2B, we extend the procedure to a more complex orthographic system, namely English.

**EXPERIMENT 1A**

The German language allows us to neatly assess the independent contributions of context-insensitive GPCs, context-sensitive GPCs and BRCs in a non-word reading paradigm: It is possible to create a set of items which generate different predictions for vowel pronunciation, depending on which strategy is used.

In German, there is relatively little ambiguity in print-to-sound correspondences, compared to English. What little ambiguity there is stems mostly from vowel pronunciation (Ziegler et al., 2000): Each vowel can be pronounced as either long or short (e.g., “Schal” /ʃa:l/ versus “Schall” /ʃa:l/). In monosyllabic words, vowel length is often signalled by context. Some CSCs allow the reader to unambiguously determine vowel length; for example, any vowel followed by an “h” is pronounced long (“V[h]” → long vowel). Other CSCs are less transparent. These correspondences are described by a German implementation of Coltheart et al.’s (2001) DRC (Ziegler et al., 2000). To allow the sublexical route to determine vowel length, it contains a set of context-sensitive super-rules: any vowel which is followed by only one consonant elicits a long vowel response (e.g., “Wal”) and a vowel which is followed by two or more consonants is pronounced short (e.g., “Wald”). These two rules can be summarised as follows: “V[C]” → long vowel and “V[C][C]” → short vowel.

Although these super-rules capture the overall statistical distribution, there are also some exceptions or words that would be irregular according to the German DRC (Ziegler et al., 2000). The word “Magd”, for example, is pronounced with a long vowel; conversely, the word “Bus” is pronounced with a short vowel. The presence of several bodies which consistently break the super-rules allows us to orthogonally manipulate the number of consonants in the body of a non-word and the pronunciation of the base words. Thus, we create a situation where the different types of correspondences (i.e., super-rules and body analogy) make different predictions about the pronunciation of the vowel.

For the present experiment, we can make a set of simple predictions if we assume that readers generally use only one type of correspondence: If only context-insensitive GPCs are used for German non-word reading, we expect that the likelihood of a short vowel pronunciation should be independent of any other orthographic features of the non-word. Such a GPC would predict many more short than long vowels, as the majority of vowels in German have the short pronunciation (Perry, Ziegler, Braun, & Zorzi, 2010). If a context-sensitive super-rule is used, vowel length should be solely determined by the number of consonants following the vowel. In this case, even a non-word based on the irregular but consistent word such as “Magd” (e.g., “blagd”) should be pronounced with a short vowel. These irregular-base word items can distinguish between reliance on super-rules compared to BRCs: if BRCs are
used, non-words based on irregular consistent words should be pronounced to rhyme with their real-word counterparts.

**Methods**

Participants were 12 German native speakers who were staff or postgraduate students at Macquarie University, or members of the university’s German society. As they lived in Australia, they were also fluent in English—a point which we will discuss in a later section. With one exception, all participants had completed secondary education in Germany and 10 had also attended German tertiary education. One participant had moved to Australia at the age of 5, but had attended a German-speaking school for seven years.

The non-words that were used for this experiment are listed in Appendix A. There were 30 non-words in each of three conditions. The non-words were created by changing the onsets of real words. All base words were taken from a list of consistent German words (J. Ziegler, personal communication, January 26, 2012). The first condition used base words with V[C] bodies which were pronounced with a long vowel (“Jod” → “FOD”); the second condition was based on V[C][C] words with a short vowel (“Saft” → “BLAFT”). The third condition was derived from irregular words, which had either a V[C] body but a short vowel (“mit” → “GIT”) or a V[C][C] structure and a long vowel (“Jagd” → “BAGD”). The three conditions were matched on orthographic N (the number of real words that can be created by substituting one letter): V[C] items had an average orthographic N-size of 1.73 (SD = 1.46), V[C][C] items had a mean of 2.10 (SD = 1.69) and items with irregular base words had a mean of 1.83 (SD = 1.90). The mean body-N (number of real words with the same body) for the three conditions is 1.93 (SD = 1.87), 2.40 (SD = 1.98) and 1.37 (SD = 1.00) respectively.

Participants were tested individually in a quiet room. Instructions were given in German by a native speaker. The participants were told that they would be asked to read non-words which were created using German orthographic rules. The instructions emphasised that accuracy was more important than speed to discourage quick lexical processing, which might result in lexicalisation errors.

The items were presented using the DMDX software package (Forster & Forster, 2003) in random order. Each trial consisted of a fixation cross, which remained in the centre of the screen for 500 ms, followed by the item, which remained on the screen until the voice key was triggered. Ten practice non-words preceded the experiment. As all nouns in German are spelled with capital initial letters, presenting non-words in all lowercase would provide an indication of word class of a non-word. Previous research has shown that information on the likely word class of a non-word affects its pronunciation (Campbell & Besner, 1981). Therefore, all items were presented in upper case.

**Results**

Six trials (0.6%) were excluded due to poor sound quality or premature voice key triggering. The rest of the trials were scored by a German native speaker as pronounced with a long vowel, a short vowel or incorrectly. For identifying incorrect responses, we used a lenient marking criterion: if a participant’s response was consistent with a possible pronunciation of the GPCs, it was marked as correct (e.g., “spic” was marked as correct regardless of whether it was pronounced as /spik/ or as /spık/—whereas, in German, “s” is typically pronounced as /ʃ/ before “p” or “t”, there are a few instances, such as loanwords, where it is assigned as the pronunciation /s/). Overall, 1.1% of all responses were classified as incorrect and excluded from subsequent analyses. Of primary interest were the proportions of long and short vowel responses and how they differed across condition. We split the irregular-base word condition by whether the bodies had one (hereafter referred to as V[C] Irregular; N = 17) or two (V[C][C] Irregular, N = 13) consonants. Note that the two “irregular” conditions did not differ dramatically on any item characteristics: the mean number of letters was 3.88 (SD = 0.34) and 4.36 (SD = 0.63) for the V[C] and V[C][C] conditions, respectively, orthographic N was 1.88 (SD = 2.19) and 1.79 (SD = 1.58), respectively, and body-N was 1.56 (SD = 1.15) and 1.29 (SD = 0.61), respectively. The proportions of short vowel responses for each of the four item types (V[C] Regular, V[C][C] Regular, V[C] Irregular and V[C][C] Irregular) are listed in Table 1, along
with the predictions according to each of the three types of correspondences.

In order to make the predictions more specific, we can use a corpus analysis to determine the percentage of times a vowel is pronounced as long or short under certain circumstances. For example, overall, 78.02% of all monosyllabic words are pronounced with a short vowel (Perry, Ziegler, Braun, & Zorzi, 2010); therefore, if German readers rely on context-insensitive GPCs, we expect them to give around the same percentage of short vowel responses. Among words with a single-consonant coda, 24.53% are pronounced with a short vowel, so we expect about the same percentage of short vowel responses to V[C] non-words, if only super-rules are used to determine vowel length. In Table 1, we present the predicted vowel lengths for each of the four conditions and by each of the three types of correspondences. For the context-insensitive GPCs and super-rules, these are calculated from the analyses presented in Perry, Ziegler, Braun, and Zorzi (2010). The predictions of the BRCs depend on the consistency ratio of the body. In the current study, we used only consistent items, where the body has only one pronunciation in real words. This means that if participants rely solely on BRCs, 100% of the pronunciations should be consistent with the base word vowel length. This is a clear indication that German readers rely on more than one type of correspondence for reading non-words. Moreover, a closer look at Table 1 shows that no combination of two types of correspondences can account for the results, either: If context-insensitive GPCs and CSCs were the sole determiners of vowel length, we would not expect to find different proportions for the V[C] Regular and V[C] Irregular items—but we do. If context-insensitive GPCs and BRCs were the only predictors of vowel length, we would find no difference between the V[C] Regular and the V[C][C] Regular items—and we do. If only CSCs and BRCs were used, we should observe less than 25% short vowel responses—which is not supported by the data.

### Modelling vowel pronunciations

It is not possible for a single or even a pair of types of correspondences to adequately fit the empirical data. It may be, however, that some combination of all three types of correspondences provides a good fit. Here we introduce a mathematical modelling approach that allows us to uncover more complex relationships between the types of correspondences. The goal is to weight the three strategies.\(^3\) More formally, we are seeking a set of \(\beta\) weights that best satisfy the following mathematical model (one pair of equations for each item):

\[ \text{Model predictions} \]

| Responses | V[C] Regular | V[C][C] Regular | V[C] Irregular | V[C][C] Irregular |
|-----------|--------------|-----------------|---------------|-----------------|
| Example   | "Wal" → "bral" | "Wald" → "brald" | "Bus" → "brus" | "Magd" → "bragd" |
| % Short 1A | 47.25        | 83.69           | 84.63         | 61.04           |
| % Short 1B | 37.28        | 86.79           | 72.95         | 62.84           |

GPC, context-insensitive GPC; CSC, context-sensitive correspondences; BRC, body-rime correspondence.

---

\(^3\) Although we refer to the reliance on different types of correspondences as a “strategy”, we do not mean to imply that readers consciously choose the type of correspondence that maximises the chance of correctly reading an unfamiliar word, in a way that optimally fits the empirical data.
where $\text{GPC}_{\text{length},i}$ is the probability of item $i$ being pronounced with a vowel of the corresponding length according to the corpus analysis when using only context-insensitive (single-letter) GPCs as a predictor, $\text{CSC}_{\text{length},i}$ is the probability according to context-sensitive super-rules and $\text{BRC}_{\text{length},i}$ is the probability according to the BRCs. Table 1 provides the average predictions for each condition, but the predictions from each correspondence were calculated separately for each item in the experiments. $P_{i}(\text{length})$ is the empirically observed proportion of the vowel length in Experiment 1A.\(^4\)

At a first glance, this would appear to be a simple regression problem (with no intercept term). Linear regression would optimally select $\beta$ values that minimised the prediction error for Equation 1 (indexed by the residual sum of squares). However, there are several reasons why this should not be thought of as regression. First, uninterpretable. This means that all of our $\beta$ parameters must exceed 0. This constraint cannot be guaranteed by standard linear regression using ordinary least squares (Monfort, 1995).

Even with only positive $\beta$s, there are two ways to interpret the weights. One could think of them as the contribution of each strategy to some sort of blending process that ultimately chooses the vowel pronunciation. In which case, we can simply fit the model in Equation 1 with the constraint that $\beta_j \geq 0, \forall i$. Alternately, one can think of the weights as the probabilities of adopting the vowel pronunciation from a given strategy. We prefer the latter interpretation (and discuss some evidence for it later), but it requires two further constraints: the $\beta$ weights must fall below 1, and, since we assume that the three strategies (GPCs, super-rules and BRCs) are exhaustive, the three $\beta$s must sum to 1. The model can be formalised as:

\[
\begin{align*}
    P_i(\text{short}) &= \beta_{\text{gpc}} \times \text{GPC}_{\text{short},i} + \beta_{\text{csc}} \times \text{CSC}_{\text{short},i} + \beta_{\text{brc}} \times \text{BRC}_{\text{short},i} \\
    P_i(\text{long}) &= \beta_{\text{gpc}} \times \text{GPC}_{\text{long},i} + \beta_{\text{csc}} \times \text{CSC}_{\text{long},i} + \beta_{\text{brc}} \times \text{BRC}_{\text{long},i}
\end{align*}
\]

(1)

\[
\begin{align*}
    P_i(\text{short}) &= \beta_{\text{gpc}} \times \text{GPC}_{\text{short},i} + \beta_{\text{csc}} \times \text{CSC}_{\text{short},i} + \beta_{\text{brc}} \times \text{BRC}_{\text{short},i} \\
    P_i(\text{long}) &= \beta_{\text{gpc}} \times \text{GPC}_{\text{long},i} + \beta_{\text{csc}} \times \text{CSC}_{\text{long},i} + \beta_{\text{brc}} \times \text{BRC}_{\text{long},i} \\
    \text{where } \beta_j &\in [0, 1] \text{ and } \Sigma \beta_j = 1, \forall j \in \{\text{gpc, csc, brc}\}
\end{align*}
\]

(2)

since the $\beta$ values are thought of as the degree to which a strategy applies in reading the items in Experiment 1, negative values would be

---

4In standard linear regression, only one of these two formulae would be required, since they are entirely independent (i.e., $P_i(\text{Long}) = 1 - P_i(\text{Short})$, etc.). In traditional regression, the only difference between the first and second equations would be the location of the estimated intercept and the sign of the slope. However, by removing the intercept term, our modelling strategy undermines this interdependence. Since the intercept is not free to vary (it is forced to be 0) the parameter estimates for $P(\text{short})$ would not match those for $P(\text{long})$. As a result, we must simultaneously fit both vowel pronunciations. Although it is useful to use the language of regression to describe some of the procedures, it is very important to remember that the $\beta$s here do not represent regression slopes but weights. Also, if this were a regression problem, it would be more properly treated as a logistic regression problem. However, this would be incompatible with our interpretation of the weights as “the probability that a certain strategy is adopted”.

---

That is, we are seeking a set of probabilistic weights on the three strategies that minimises the prediction error of the model. The challenge here is to both efficiently search the available parameter space and satisfy the $\Sigma \beta_j = 1$ constraint. The first problem is a well-studied one in computer science and solutions are available that solve it. The second problem is largely solved by introducing an additional equation that can only be satisfied if $\Sigma \beta_j = 1$, and giving that equation a strong influence on the final parameter set. The interested reader can find a fuller discussion of the implementation details in Appendix B.

**Optimal weights in Experiment 1A**

In Experiment 1A, we collected the proportion of short and long vowel responses to 90 items, and
for each item we have the predicted probability of a short or long vowel pronunciation according to each of the three strategies. The strategy predictions were obtained from the corpus analysis undertaken by Perry, Ziegler, Braun, and Zorzi (2010). Using the technique described earlier, the native German readers in Experiment 1A appear to be relying mostly on GPCs ($\beta_{gpc} = 0.56$), and to a lesser extent on super-rules ($\beta_{csc} = 0.19$) and BRCs ($\beta_{brc} = 0.26$). See Table 2 for a summary of the modelling results across all of the present experiments.

The previous analysis contains a theoretically supported but strong assumption that readers use only the three strategies described in the introduction when reading non-words. It is possible that other sources of information are used by German native speakers to determine vowel length. We can provide a simple test of this possibility by relaxing some of the constraints on the model and observing how critical those constraints were to the optimisation results. To do this, we removed the $\Sigma \beta_j = 1$ constraint and allowed the $\beta$ s to take on any positive weights in the fitting process. That is, we fit the following alternative model (some subscripts indicating length and item have been omitted for simplicity):

$$P(\text{Length}) = \beta_{gpc} \times \text{GPC} + \beta_{csc} \times \text{CSC} + \beta_{brc} \times \text{BRC}, \text{ where } \beta_i > 0 \ \forall i$$

If readers are adopting other strategies that are not well described by the GPC, super-rules and BRC strategies, the incomplete nature of the model should be reflected in these alternate weights. The weights that optimise Equation (3) were $\beta_{gpc} = 0.58$, $\beta_{csc} = 0.14$, and $\beta_{brc} = 0.24$. These values sum to 0.96, suggesting that there is little need for a fourth strategy to describe the data. This does not conclusively rule out a role for any other strategies, but provides some evidence that the three strategies already tested are sufficient. That said, there is one additional strategy that could be playing a role: anti-body correspondences (ABC) or the probability of a vowel being pronounced as long or short based on the onset of the word. In this corpus of non-words, the predictions from ABC and context-insensitive GPCs are highly correlated, so it is difficult to disentangle the two strategies entirely, but it may be that ABC are more important than context-insensitive GPCs and thus are a better predictor. To test whether or not ABCs were important for determining vowel pronunciations, we added a component to model (2):

$$P(\text{Length}) = \beta_{gpc} \times \text{GPC} + \beta_{csc} \times \text{CSC} + \beta_{brc} \times \text{BRC} + \beta_{abc} \times \text{ABC}$$

where the addition of ABC represents the predictions from ABCs, and $\beta_{abc}$ is the associated weight. Fitting Equation 4 produced the same weights that resulted from Equation 2 where the ABCs were not included. That is, $\beta_{abc} = 0$, giving little reason to believe that any other strategies are being used in Experiment 1A.

**Model fits.** The optimisation procedure presented here is only useful if it arrives at a model that fits the data better than alternatives. To determine the effectiveness of the model, we calculated the goodness of fit for each model using the correlation coefficient. The model that resulted from fitting Equation 3 was selected as the optimal model for Experiment 1A. Table 3 presents a summary of how well the model fits the data for each of the four experiments.

| Sample | GPC | CSC | BRC | Optimal (95% CI) |
|--------|-----|-----|-----|------------------|
| 1A (German bilingual) | 0.714 | 0.681 | 0.540 | 0.844 (0.830, 0.847) |
| 1B (German monolingual) | 0.578 | 0.730 | 0.659 | 0.827 (0.812, 0.832) |
| 2A (English monolingual) | 0.522 | 0.630 | 0.385 | 0.729 (0.719, 0.731) |
| 2B (English bilingual) | 0.514 | 0.573 | 0.568 | 0.792 (0.785, 0.793) |

### Table 2

| Correspondence type | 1A (German bilingual) | 1B (German monolingual) | 2A (English monolingual) | 2B (English bilingual) |
|---------------------|-----------------------|------------------------|-------------------------|-----------------------|
| GPC                 | 0.56                  | 0.38                   | 0.05                    | 0.03                  |
| CSC                 | 0.19                  | 0.35                   | 0.69                    | 0.61                  |
| BRC                 | 0.26                  | 0.27                   | 0.26                    | 0.36                  |
correlation between the model predictions and the observed response patterns. For comparison, we did the same for the GPCs, CSCs and BRCs individually. As can be seen in Table 3, the optimisation process outperforms the other three alternatives in all four samples presented here. In Experiment 1A, the correlation is .844, whereas the next best model (based on context-insensitive GPCs) correlates at .714.

Discussion

In Experiment 1A, we successfully used an optimisation procedure to quantify the degree of reliance on three types of sublexical correspondences: context-insensitive GPCs, context-sensitive GPCs and BRCs. This can be achieved with the German language because it is possible to create items where different correspondence types make different predictions about the vowel length pronunciation.

Importantly, we found that all three types of correspondences are both necessary and sufficient to predict vowel length responses in a sample of German native speakers. Context-insensitive correspondences appear to be the strongest predictor. This is in line with the psycholinguistic grain-size theory, which argues that the smallest unit size is favoured by readers of a language with predictable GPCs, such as German (Ziegler & Goswami, 2005).

Experiment 1A has some limitations. It could be argued that the results are unreliable, first due to the small sample size and second because the participants were bilingual, and very fluent in English. It is unclear how fluency in English may affect the reliance on different types of correspondences in German. Even though we took care to only include German participants who learned to read and write in German from a young age, there is a possibility that their exposure to German reading material has been diminished by residing in an English-speaking country. It is also possible that their knowledge of English would change the preferred unit in their native language: for example, psycholinguistic grain-size theory predicts that readers of English rely more heavily on larger grain sizes than readers of German (Ziegler & Goswami, 2005), although it does not make any statements about sublexical processing in bilinguals. We address these concerns in Experiment 1B.

EXPERIMENT 1B

In Experiment 1B, we collected data with two different samples of German native speakers who live in Germany and are not exposed to English on an everyday basis. We hereafter refer to them as monolingual Germans, even though they are not strictly monolingual: due to globalisation, it would be difficult if not impossible to find Germans who have no knowledge of English. Having collected data with two different samples of monolingual Germans allows us to test the reliability of the modelling method described here. If our model arrives at similar weights for two independent samples from the same population, we can be more confident that our modelling procedure is stable and reliable.

Methods

The methods were almost identical to Experiment 1A. One item was replaced (due to a typo, the original item set contained an inconsistent item, “blen”, which was changed to “blem” in Experiment 1B).

The first sample consisted of 10 German native speakers who were staff or students at the Freie Universität in Berlin. All had completed their schooling in Germany. The second sample consisted of 26 undergraduate students at Potsdam University. Again, all were native German speakers and had completed their education in Germany.

Results

The scoring procedure was identical to Experiment 1A. For the Berlin sample, there were two non-responses (0.22%) and 15 errors (1.67%). The Potsdam sample made 2.3% errors. A series of $t$-tests showed that the percentages of long and short vowel responses did not differ significantly for any of the conditions across the two samples, all $p > 0.4$. Furthermore, fitting each sample separately using the model described in Equation 2 produced very similar weights. For the participants from Berlin, the weights were $\hat{\beta}_{\text{gpc}} = 0.40, \hat{\beta}_{\text{csc}} = 0.33$, and $\hat{\beta}_{\text{brc}} = 0.27$. For the participants from Potsdam they were $\hat{\beta}_{\text{gpc}} = 0.37, \hat{\beta}_{\text{csc}} = 0.35$, and $\hat{\beta}_{\text{brc}} = 0.28$. This result is comforting, suggesting that the method introduced here is reliable across different samples from similar populations. Since there was
little difference between the two samples, we collapsed across them yielding a sample of 36 native German monolinguals. Using this collapsed sample, our model produces $\hat{\beta}_{\text{gpc}} = 0.38$, $\hat{\beta}_{\text{csc}} = 0.35$, and $\hat{\beta}_{\text{hec}} = 0.27$. As in Experiment 1A, the optimal parameter set outperforms the alternatives in fitting the observed data (Table 3).

**German/English bilingual VS. German monolingual readers.** Since Experiments 1A and B are based on the same set of items, we have the opportunity to compare how the bilingual readers differed from the monolingual readers. The critical question is whether or not the smaller $\hat{\beta}_{\text{gpc}}$ and larger $\hat{\beta}_{\text{csc}}$ for monolinguals represents a real difference, or simply random variation. In the usual context of a linear regression model, this would be a simple matter of including the language status of the participants (bilingual vs. monolingual) in the model, and testing for an interaction between language status and the GPC and/or CSC estimates. However, our modelling strategy violates many of the assumptions that allow for straightforward $t$-tests of the parameter estimates (given the constraints of our model, the parameter estimates are unlikely to be well behaved, statistically). Instead, we turn to a bootstrapping methodology to allow us to use the data to conduct non-parametric tests of the variability in our estimates.

To establish the reliability of the difference in the $\hat{\beta}_{\text{gpc}}$ and $\hat{\beta}_{\text{csc}}$ estimates, we repeatedly resampled 90 items (with replacement) from the data-set, and estimated the $\beta_i$ s for both the bilingual and monolingual participants with each sample of items. Of 10,000 such samples, 9890 (98.9%) produced a larger GPC weight for the bilingual subjects than for the monolingual subjects (95% CI of the difference: 0.019–0.327). Similarly, 9634 (96.5%) samples produced a larger CSC weight for the monolingual participants than for the bilingual participants (95% CI: −0.011 to 0.317). This suggests that the difference in the GPC weights is robust, whereas the difference in the CSC weights is slightly more tenuous. The difference in the BRC weights was not at all significant: 3454 (34.5%) of the samples produced larger BRC weights for bilinguals than for monolinguals (95% CI: −0.058 to 0.089). We also took advantage of these bootstrap samples to estimate the variability in the correlations from the optimal parameters in Table 3.

To summarise the results so far, the reliance on BRCs did not differ between monolingual and bilingual readers, but there was a very stable difference in the reliance on context-insensitive GPCs and a somewhat stable difference in the role of context-sensitive super-rules. Monolinguals relied less on context-insensitive GPCs and somewhat more on super-rules than bilinguals.

**Individual differences.** There is some ambiguity in interpreting the weights: as we collapsed across participants, the weightings do not give us any information about inter-individual participant variability. Theoretically, it is possible that all participants rely on the same strategies to the same extent, or that the weightings are reflective of the percentage of participants who rely on a particular strategy only. To address this, we generated the weightings for each individual participant in Experiments 1A and B. These are summarised in Figure 1. This figure shows that there is individual variability, but most participants rely on a combination of the three strategies.

**Discussion**

As in the previous experiment, we were able to quantify the degree of reliance on each of the three types of correspondences in two samples of monolingual German native speakers. Even though there is individual variation, we found, on average, almost identical reliance on the three strategies in two independent samples of German readers, suggesting that the procedure we introduced is reliable. The overall pattern of results was also broadly consistent with the findings from Experiment 1A, showing that reliance on all three types of correspondences is both necessary and sufficient to explain the vowel length pronunciations in German, and that context-insensitive correspondences are the major predictor of the vowel responses.

Although the bilingual and monolingual participants’ response patterns were similar, we did find some significant differences in terms of reliance on context-sensitive versus context-insensitive correspondences: bilingual participants show stronger reliance on context-insensitive correspondences and less reliance on CSCs. Two possible causes of the difference between German/English bilinguals and German monolinguals are the influence of English proficiency on reading in the bilingual sample, or a general difference in German reading proficiency. According to the psycholinguistic grain-size theory, if the difference in weights is
due to the influence of English (L2) on the choice of correspondences in German (L1), we would expect bilinguals to rely more on larger correspondences (CSCs or BRCs as opposed to context-insensitive correspondences). Developmental studies have shown that reliance on larger units differs as a function of reading efficiency, as younger children rely to a greater extent on context-sensitive rules (Treiman et al., 2006). In Experiment 1B, we found that bilingual participants rely more on context-insensitive rules, which is more in line with a proficiency explanation—bilinguals may be less proficient in reading German than monolinguals, as they are less exposed to German texts. As a result, they rely to a greater extent on the context-insensitive correspondences.5

EXPERIMENT 2A

The majority of prior research on the use of GPCs, context-sensitivity and BRCs has been conducted in English. In contrast to German, the English letter-to-sound correspondence system is highly complex, as a large set of correspondences on different levels are required to describe the relationship between print and speech (Venezky, 1970). In Experiment 2, we aimed to explore whether it is possible to apply the methodology which we introduced in Experiment 1 to quantify the degree of reliance on the same three strategies in a more complex system.

English, like German, contains some CSCs. However, there are no super-rules, or correspondences which apply to all vowels, as in German. Therefore, we concentrated solely on the grapheme “a”, as its correct pronunciation can often be disambiguated by taking into account its context. By default, “a” is pronounced as in “cat” in Australian English, but there are several context-sensitive and multi-letter GPCs that can modify its pronunciation. The context-sensitive correspondence of interest here is the correspondence that an “a” preceded by a “qu” or “w” is pronounced as /ɔ/. We chose this correspondence to assess reliance

Figure 1. Weightings for each individual participant, sorted by degree of reliance on GPC. GPC = context-insensitive rules, CSR = context-sensitive rules, AN = body-rime correspondences. M = monolingual participant (Experiment 1B), B = bilingual participant (Experiment 1A)
on context-sensitivity for two reasons: First, previous research has shown that there is some psychological reality to this correspondence (Patterson & Behrmann, 1997; Treiman et al., 2003). Second, unlike other context-sensitive GPCs (e.g., “a[l]” → /ɔ/), this correspondence is not confounded with body–rime analogy, as the modifier is located in the onset, before the vowel. This is therefore one of the few English CSCs that allows us to independently assess effects of context-sensitivity.

In order to create an item set equivalent to the German non-words used in Experiment 1, we isolated English bodies with the vowel grapheme “a” which are consistently pronounced irregularly (Ziegler et al., 1997). There are five such bodies: “-alse”, “-att”, “-alk”, “-alt” and “-ald”. With one exception, they are confounded with the “a[l]” → /ɔ/ correspondence: the body “-att” only occurs in the word “watt” and therefore only has the /ɔ/ pronunciation. As a result, and in contrast to the German experiment, the degree of reliance on BRCs cannot be assessed using this paradigm because it is almost perfectly confounded with reliance on the “a[l]” context-sensitive correspondence.

In short, there are three possible pronunciations indicative of reliance on different types of correspondences. If English participants rely on context-insensitive GPCs, we should find that the majority of non-words are pronounced with the /æ/ vowel. If CSCs are used, then in the conditions where a “qu” or “w” precedes the vowel we should find many /ɔ/ responses. If either BRCs or the “a[l]” correspondence are used, the conditions with the consistently irregular bodies should be pronounced with an /ɔ/.

Methods

The participants were 19 undergraduate students at Macquarie University who were all native speakers of English.

We created four conditions of 18 words each (listed in Appendix A). All were monosyllables containing the single vowel grapheme “a”. The first condition was created by taking consistently regular bodies (Ziegler et al., 1997) and adding an onset which does not change the pronunciation of the vowel (i.e., any onset that does not contain “w” or “qu”), resulting in non-words like “hact” (this condition is hereafter referred to as CS+BR+). The second condition was created by taking the single vowel grapheme “a” (this condition is hereafter referred to as CS+BR-). The second condition (CS+BR-, e.g., “halse”) was based on bodies where the “a” is consistently pronounced as /ɔ/ (or /ɔ/ for the body “-att”) and “normal” onsets, as in the first condition. Here, the BRCs predict an /ɔ/ pronunciation, and therefore disagree with the context-insensitive correspondence. The items in the third condition (CS–BR+, e.g., “wact”) were based on regular bodies and onsets containing “w” or “qu”, meaning that the context-sensitive “[qu,w]”a-correspondence contradicted the context-insensitive GPC, whereas the BRCs did not. The fourth condition (CS–BR–, e.g., “quals”) had items with irregular bodies and onsets with “w” or “qu”—here both the context-sensitive correspondence and the body disagree with the context-insensitive GPC. As filler items, we used a set of unrelated non-words.

The presentation was identical to Experiment 1, with items presented in random order and in upper case letters. As with Experiment 1, participants were instructed to read the items as accurately as possible, without putting them under time pressure.

Results

The results were scored by the fourth author (SP), a native Australian English speaker and an experienced transcriber, with the aid of spectral analysis using the EMU speech database system and associated speech analysis tools (Cassidy & Harrington, 2001). SP was unaware of the aims of the experiment as she was transcribing the data. Unlike the German data, scoring the responses as correct or incorrect was more complicated. For the grapheme “a”, there are at least five plausible pronunciations: as in “cat”, as in “false”, as in “what”, as in “cake” and as in “car”. We considered only the first three responses, as they were predicted either by the context-insensitive GPC, “a” → /æ/, the context-sensitive GPC, “[qu,w]a” → /ɔ/ or the BRC “a[l]” → /ɔ/ context-sensitive correspondence. Other responses and errors made up 4.09% of the CS+BR+ condition, 24.85% of the CS+BR– condition, 6.43% of the CS–BR+ condition and 20.76% of the CS–BR– condition, and were excluded from the subsequent analyses. The percentage of “other” responses is particularly high for the BR– conditions, partly because in English, a post-vocalic “l” creates ambiguity in the pronunciation of the vowel, such that a long /ɔ/ may become indistinguishable from the phoneme /æ/. The percentages of /æ/, /ɔ/ and /ɔ/ responses are presented in Table 4, with the results from Experiment 2B for comparison.
Modelling vowel pronunciations in English

The modelling strategy for Experiments 2A and B required a small modification from that employed in Experiments 1A and B. In German, there are only two available vowel pronunciations for “a”: short and long. In Australian English, there are three pronunciations available for items of Experiment 2. This means that we now need three equations per item (item indices are omitted):

\[
P(\text{æ}) = \beta_{\text{gpc}} \times \text{GPC}_{\text{æ}} + \beta_{\text{csc}} \times \text{CSC}_{\text{æ}} + \beta_{\text{brc}} \times \text{BRC}_{\text{æ}}
\]

\[
P(\text{ɔ}) = \beta_{\text{gpc}} \times \text{GPC}_{\text{ɔ}} + \beta_{\text{csc}} \times \text{CSC}_{\text{ɔ}} + \beta_{\text{brc}} \times \text{BRC}_{\text{ɔ}}
\]

\[
P(\text{o:}) = \beta_{\text{gpc}} \times \text{GPC}_{\text{o:}} + \beta_{\text{csc}} \times \text{CSC}_{\text{o:}} + \beta_{\text{brc}} \times \text{BRC}_{\text{o:}}
\]

where each of the subscripted strategies indicates the likelihood of the subscripted pronunciation under that strategy; for example, GPC_{æ} indicates the likelihood of an /æ/ response under the GPC strategy. The end result is a set of \( \beta_j \)s that fit all three pronunciations simultaneously.

The weightings are shown in Table 2. The role of CSCs appears to be the most important in predicting the pronunciation of the grapheme “a”, with \( \beta_{\text{csc}} = 0.69 \). BRCs also appear to contribute significantly, \( \beta_{\text{brc}} = 0.26 \), whereas the reliance on context-insensitive correspondences is very small, \( \beta_{\text{gpc}} = 0.05 \). Indeed, the bootstrapping procedure produced \( \beta_{\text{gpc}} = 0 \) in 43.3% of the samples and \( \beta_{\text{gpc}} < 0 \) in 82.0% of the samples, suggesting that the reliance on context-insensitive correspondences does not differ significantly from zero. Here again, the model is outperforming each of the independent strategies at predicting response patterns on an item by item basis (see Table 3), but when considering the model’s ability to predict cell means (Table 4), it is clear this approach is less successful in English than it was in German.

Discussion

We quantified the reliance on different types of correspondences for English non-words with the grapheme “a”, using the same modelling technique we introduced in Experiment 1 for German, with...
some minor modifications. Although the results were less clear-cut than in German, we show that the procedure can be applied to a more complex orthography. The model fits in Table 4 indicate that the English orthography is not best suited for such an analysis. In particular, the poor model fits are due to many /ɔ:/ responses, even when these were not predicted by our model. This may be a result of the complex phonology of English: the phonemes /ɔ/ and /o:/ are very similar, therefore it is possible that the participants had a tendency to shorten /ɔ:/ responses, which then became indistinguishable from the vowel /ɔ/. The second possibility is that another source of information is used to determine vowel pronunciations in English which we did not take into account.

Despite these limitations, there are several conclusions that can be drawn from the results. First, the weightings showed that in English the three strategies are neither necessary nor sufficient to predict the pronunciation of the grapheme “a”. In contrast to German, we obtained a relatively high percentage of “other” responses for the English data, or pronunciations that were implausible according to any of the correspondences that we thought participants may use. Such a heterogeneity of non-word reading aloud responses has also been reported elsewhere (Andrews & Scarlett, 1998; Pritchard et al., 2012). Although this would be an interesting topic to pursue in further research, for our purposes we discarded the unusual pronunciations as we were interested in quantifying the reliance on the same three types of correspondences we showed to be critical to non-word reading in German. This high percentage of “other” responses shows that it is likely that other strategies, such as more complex CSCs or lexical analogy, are used during non-word reading in English. In other words, the three types of correspondences we described in the introduction are not sufficient to explain vowel responses to the grapheme “a” in English—which is in contrast to the findings we report for German.

Second, a striking finding is that the context-insensitive correspondences are hardly used at all to derive the pronunciation of the grapheme “a”. Rather, English readers rely heavily on the context-sensitive GPC, which can often be used to derive the correct pronunciation for English words.

These results imply that in the special case of the grapheme “a”, it may not be necessary to rely on all three types of sublexical correspondences to explain the pattern of vowel responses. We consider it highly unlikely that context-insensitive GPCs are not used at all for reading in English. We relied solely on non-words with the grapheme “a” to derive the weightings in Experiment 2, and its correct pronunciation can often be predicted by context. Arguably, this may falsely bias the weightings towards an apparent greater reliance on CSCs than we would observe if we used different graphemes for this procedure. However, we consider it likely that context-sensitivity plays an equally important role for other vowels in English: as is the case for the grapheme “a”, vowel pronunciations in English are generally inconsistent, but can be often resolved CSCs (Treiman et al., 1995). Non-word reading studies have also provided evidence for the psychological reality of CSCs determining vowel pronunciation in English, other than the “[qu/w]a” correspondence (Treiman et al., 2003, 2006). As described earlier, we focused on the “[qu/w]a” correspondence only because it is not confounded with BRCs—if we used any other context-sensitive correspondence we would be unable to distinguish it from reliance on body analogy.

Again, we stress that the almost exclusive reliance on CSCs in Experiment 2 is unlikely to generalise to the processing of more consistent graphemes in English, such as consonants. If, linguistically, context-insensitive correspondences are generally predictive of the correct pronunciation, there is no pressure on the readers to take into account the surrounding letters for those particular graphemes.

As discussed in the Introduction, the BRCs of English are confounded with CSCs. Instead of the German super-rules, we used an English context-sensitive correspondence that is not located in the body, namely the “[qu,w]a” → /ɔ/ correspondence. However, we cannot fully disambiguate the reliance on BRCs and the “[ɔl]” correspondence. Future studies using non-word reading should bear in mind that BRC and CSCs are heavily confounded, and that an apparently irregular pronunciation of a non-word may show reliance on either CSCs or BRCs.

**EXPERIMENT 2B**

In Experiment 2B, we tested a sample of German/English bilingual speakers on the English item set. As with Experiment 1B, this will allow us to verify the weightings in a different sample and explore...
potential differences between mono- and bilingual participants.

In Experiment 1, we argued that the differences that we found between the two samples are more consistent with an account based on reading proficiency rather than one based on the influence of acquiring a language with a deeper orthography. However, it may be that an early acquired L1 shapes the cognitive system in a way that biases the processing of subsequently learnt languages towards familiar types of correspondences. If so, this would predict a difference between participants reading English non-words depending on whether their first language was English (as in Experiment 2A) or German.

Methods

The participants were 13 native German speakers living in Australia (undergraduate and graduate students at Macquarie University, academic staff, family and friends). Eight of them had also participated in Experiment 1A several months earlier, but did not know that the two studies were related. In this sample, all participants had lived in Germany for at least 18 years before moving to an English-speaking country. The items and procedure were identical to Experiment 2A. The participants were told that they would see English non-words and were asked to pronounce each item as if it were an English word that they are unfamiliar with.

Results

The same scoring system was used as for Experiment 2A. The proportions of /æ/, /ɔ/ and /o:/ responses for both Experiments 2A and B are presented in Table 4. German native speakers overall gave more “other” non-word responses or vowel responses that were inconsistent with our predictors, compared to the English monolinguals in Experiment 2A: 15.74%, 23.61%, 17.95% and 8.80% for the CS+BR+, CS+BR–, CS–BR+ and CS–BR– conditions, respectively.

We repeated the optimisation technique to derive strategy weights for this Experiment. Table 2 summarises the weights for each of the three strategies in Experiments 1A and B and 2A and B. The results of Experiment 2B mirror the findings from Experiment 2A: Again, we find strongest reliance on CSCs, robust reliance on BRCs, and negligible reliance on context-insensitive correspondences. Numerically, the reliance on CSCs appears to be larger (\( \hat{\beta}_{\text{csc}} = 0.61 \)) than in the monolingual sample (\( \hat{\beta}_{\text{csc}} = 0.69 \)). Here again, the optimal parameters outperform the alternatives with a correlation of .717 (see Table 3).

Comparing bilingual to monolingual English readers. Using the same bootstrapping technique described in Experiment 1, we confirmed that the German-English bilingual participants relied more on BRCs (BRCs) than did the English monolinguals. In 9998 (99.98%) of the samples, \( \hat{\beta}_{\text{brc}} \) was larger for bilinguals than monolinguals (95% CI of the difference: 0.046–0.150). The two samples did not differ significantly in their reliance on context-insensitive (GPC) rules, but there is some evidence that the monolinguals may rely more on CSCs (91.72% of the samples, 95% CI: −0.039 to 0.160).

Discussion

In Experiment 2B, we collected data on English non-word pronunciation from German/English bilingual participants, which we then compared to the “a”-pronunciations of English monolinguals in Experiment 2A. Again, we find that the fits of the model are somewhat discrepant with the data, suggesting that the pronunciation of the letter “a” depends also on sources of information that are not included in our model. As in Experiment 2A, we found no reliance on context-insensitive GPCs in either group, and only a non-significant trend towards larger reliance on BRCs or the “a[ll]” \( \rightarrow /o:/ \) correspondence in English monolinguals than the German/English bilinguals.

We found broadly the same pattern among two different groups of participants; here, we once again demonstrate the reliability of the optimisation procedure. The significant difference in the reliance on BRCs suggest that German native speakers, when they are highly proficient in English, rely more on these large units than English monolingual participants. Thus, the native orthography does not appear to leave noticeable footprints in the cognitive processes underlying reading in a second language, as in that case we would expect diminished reliance on BRC in German compared to English native speakers.
GENERAL DISCUSSION

In four experiments, we explored the reliance on three different sublexical correspondence types in different populations. In Experiments 1A and B, we found that German native speakers relied on all three strategies: the greatest weighting was found for context-insensitive GPCs, followed by context-sensitive GPCs (super-rules) and BRCs when reading German-derived non-words. In Experiments 2A and B, we applied the same procedure to quantify the types of correspondences that participants rely on to derive the pronunciation of the grapheme “a” in English. We found strong reliance on context-sensitive GPCs, some reliance on BRCs and little evidence that context-insensitive GPCs play a large role in determining the pronunciation of the grapheme “a”.

Cross-linguistic differences in the choice of sublexical correspondences: comparing Experiments 1 and 2

Previous theoretical work predicts cross-linguistic differences in the reliance on different units in German and English (Ziegler & Goswami, 2005). Unfortunately, with the experiments in the current study it is impossible to make a direct quantitative comparison across the two languages as we are comparing two differently structured orthographic correspondences. An alternative approach is to conduct the analyses within the languages and point out the differences between them on a descriptive level.

Our data suggest that given a grapheme where context is very important in English (i.e., “a”), context-sensitivity becomes very important compared to German, where context-insensitive correspondences are the major predictor. This is true even for a situation where there are statistical regularities at the level of CSCs. This is broadly in line with the psycholinguistic grain-size theory (Ziegler & Goswami, 2005): as the context is often an important predictor of the correct pronunciation of English words, readers are forced to rely on larger units. Our data emphasises the importance of context-sensitive GPCs in an inconsistent orthography such as English. In German, on the other hand, context-insensitive correspondences are mostly sufficient to derive the correct pronunciation of an unfamiliar word; therefore, this level of correspondences is preferred.

The reality of the cross-linguistic differences becomes more evident in a comparison of Experiments 1A and 2B. This is partly a within-subject design and involves bilingual participants reading both the English and the German item sets. The differences between the weightings in these two experiments were remarkable, with the pattern of results being more similar to that of the monolinguals of the respective language. This shows that the language is the determining factor for the reliance on different unit sizes, rather than the language background of the participants.

From this comparison, we conclude that the language that a participant is asked to read in matters more than the participant’s language background: comparing the participants in Experiments 1A and 2B shows that bilinguals rely on the three types of correspondences almost to the same extent as monolinguals do in their respective language. Thus, we conclude that the cross-linguistic differences in sublexical processing are language-specific: acquiring a deep versus shallow orthography from childhood does not shape the cognitive system, but rather encourages the reader to rely on certain types of correspondences above others in that particular orthography. Those preferences do not seem directly transferable to a later acquired orthography; instead, a reader develops a sensitivity to the most advantageous combination of strategies in the new language.

Models of reading

The current study shows that both in English and in German, several correspondence types are used in parallel. There are multiple verbal models that postulate such a scenario (LaBerge & Samuels, 1974; Patterson & Morton, 1985; Taft, 1991, 1994; Ziegler & Goswami, 2005). The theoretical contribution of the current paper is proposing a method to quantify the degree to which these are used, which can be used as a benchmark for computational models.

An open question then is whether the current computational models can simulate the obtained results. The parallel processing of various correspondences poses a computational problem: whenever there are conflicts between the pronunciations predicted by various correspondences, the system needs a way to resolve these. In English, this is important because there are often cases where different sublexical correspondences provide conflicting information.
In Table 5, we provide the percentages of regular responses from two models which have been implemented both in English and in German, namely the DRC (Coltheart et al., 2001; Ziegler et al., 2000) and the connectionist dual process (CDP+) model (Perry, Ziegler, Braun, & Zorzi, 2010; Perry, Ziegler, & Zorzi, 2007). For English, there is a newer version of the CDP+, namely the CDP++ (Perry, Ziegler, & Zorzi, 2010), which differs from the CDP+ in several points: it has been trained on a larger word set, contains some parameter changes and can also deal with polysyllabic words. We provide the simulation data from both versions of the model.

Both the CDP+/CDP++ and the DRC are dual route models of reading, where non-words are read purely via a sublexical procedure. Therefore, the current data are relevant to both models, as it concerns the nature of sublexical processing. The distinguishing feature between the two models is the way in which this procedure operates. The DRC has a set of sublexical GPCs, which are manually programmed into the sublexical route. A GPC in the DRC is defined as the most frequent phoneme that co-occurs with a given grapheme. As described in the Introduction, the DRC contains CSCs as well as single-letter and multi-letter correspondences, but there is some ambiguity when it comes to deciding which CSCs to include in the model. The current version of the English DRC does not contain either a “[w]a” or an “[a][l]” correspondence; therefore, it provides the response /æ/ to all items (see Table 5). For the second DRC simulation, we added some more CSCs; however, this does not seem to reflect the overall responses given by participants, either, as it now underestimates the number of regular (i.e., /æ/) pronunciations given by the participants. For the German DRC, the GPCs that are used to determine vowel length are the super-rules (Ziegler et al., 2000). It is clear, both from the present study (see Table 5) and from Perry, Ziegler, Braun, and Zorzi (2010) that the super-rules are not sufficient to explain German non-word pronunciations.

The CDP+/CDP++, like the DRC, is grapheme-based, but it develops context-sensitivity because the GPCs are derived via a learning algorithm, which uses real word knowledge to obtain the most likely correspondences between print and speech (Zorzi, 2010). Yet, the CDP+ does not provide an optimal fit for either the German or the English data, as it often underestimates the number of regular pronunciations (see Table 5). In particular, the English CDP+ and CDP++ seem to take CSCs into account more than the participants do, as they underestimates the number of /æ/ responses for the CS- conditions. In German, the biggest discrepancy between the CDP+ prediction and the behavioural data is in the BR- conditions, suggesting that CDP+ does not develop the same degree of reliance on BRCs that participants do.

As neither of the computational models is compatible with the behavioural results, these data cannot be used to adjudicate between the DRC and CDP+ approach. (Note that this was not the aim of the study to begin with.) We therefore turn to verbal models to provide a theoretical framework that can explain our obtained results. One such model which provides a means for the cognitive system to resolve conflicts between different sublexical correspondences has been proposed by Taft (1991, 1994). This interactive activation model states that activation passes hierarchically from the smallest units, through subsyllabic and syllabic units and morphemes to whole words, which then gives access to the semantic concept. There are additional feedback connections, which send activation from larger to smaller units.

Taft’s (1994) model also makes some explicit statements about cross-linguistic differences: the salient sublexical correspondences differ depending on the orthographic and phonological properties of the language. For example, whereas English readers parse words into orthographic-syllabic units called BOSsEs (Taft, 1979, 1992), French readers rely more on the phonological syllable (Taft & Radeau, 1995). In our experiments, we...
found reliance on similar types of correspondences in English and German. Thus, the correspondences that have psychological reality in English and German appear to be very similar. It is noteworthy that English and German are very similar in terms of their phonological and orthographic structure; therefore, we expect that the salient sublexical correspondences do not differ greatly. The situation might be different in other languages. For example, when there is a tendency for words to be polysyllabic and to contain fewer consonant clusters, as is the case in languages like Italian, Spanish or Russian, BRCs are unlikely to play a large role in reading (Duncan et al., 2013; Kerek & Niemi, 2012).

**Limitations and future directions**

The goal of the study was to identify an optimal combination of different sources of information in deciding which vowel pronunciation is most appropriate when there are two or more alternatives. A limitation of the model is that it makes no claims about the decision-making mechanisms that resolve the ambiguity, only that some sources of information are more influential than others. It may be that on each trial, the decision is based on a “winning strategy” in which the weights represent the likelihood of a particular strategy winning. Alternately, it may be that all three sources of information are combined in a Bayesian sense of “what response is most likely correct given the mix of influences”. In this case, the model weights should be interpreted as the degree of influence that each strategy has on the decision process. The present study is not able to adjudicate between these alternatives (or any others that we may not have considered), so we refrain from making strong statements favouring one or the other. The extent to which non-word pronunciations remain stable in different situations, the factors that influence any variability and the mechanisms that resolve ambiguity remain questions for future research. We do note, however, that as there is considerably variability between subjects in terms of their strategy weights (see Figure 1), there is some recent evidence that readers can be grouped according to their choices (Robidoux & Pritchard, 2014), so there may be more structure hiding within this variability.

A limitation of the paradigm as described in this paper is that it is better suited for across-subject comparisons than across-item comparisons, due to the small number of available items. This is a general problem with this approach: There are not many items where CSCs and BRCs can be dissociated, as these are intrinsically correlated. Although it would be interesting to use the same paradigm for a different set of non-word or word items to explore systematic changes in the weightings associated with item characteristics such as frequency (for words) or word-likeness (as measured, e.g., by orthographic $N$), the small number of possible items prevents us from doing this in a meaningful way.

Arguably, the data reported in this paper are also limited by our focus on the grapheme “a” only. As this criticism applies to the English data, the German data can be generalised to predicting vowel length across different graphemes. The English results and our conclusions based on these analyses are therefore weaker than those from the German analyses. Nevertheless, understanding the principles underlying reading in languages other than English is essential for the long-term goal of describing all differences and similarities between reading in different languages, and thereby creating a universal model of reading (Frost et al., 2012). This is especially important given the focus of previous literature on English. English is considered to be an “outlier” orthography; therefore, it is questionable to use it as a base for most models of skilled reading, reading development and dyslexia (Share, 2008). Although we acknowledge that, in the current context, the optimisation procedure works better for German than English, we argue that the English data provides a strong demonstration of the parallel use of different types of sublexical grain sizes, and in particular CSCs in English, new insights into cross-linguistic differences associated with the reliability of print-to-speech correspondences and a new benchmark for computational models of reading aloud.

We believe that this approach also has some utility when applied to other areas of psycholinguistics. In future research, the same paradigm can be used to systematically explore the sources of individual differences that we report in the current study. The paradigm can also be used with children: previous literature has debated for decades whether children start learning to read using large or small units first (Goswami, 2002; Goswami & Bryant, 1990; Hulme et al., 2002). Such explorations in group and individual differences are of theoretical and practical value. Future research can also apply the same mathematical procedure to any situation in which items can be
created where different strategies yield different predictions. Other areas in psycholinguistics to which this paradigm can be extended could be topics such as stress assignment for polysyllabic words, because it has been shown that, in several languages, different cues are used by participants to determine the stress of a given non-word (Arciuli, Monaghan, & Seva, 2010; Burani & Arduino, 2004; Protopapas, Gerakaki, & Alexandri, 2006; Seva, Monaghan, & Arciuli, 2009).

Conclusions

The current study contributes to the literature on cognitive processes underlying reading in several aspects. We show that context-insensitive GPCs, super-rules and BRCs are necessary and sufficient to explain the vowel length pronunciations in German; in English, context-insensitive GPCs play a smaller or negligible role in assigning the pronunciation of the grapheme “a”. We introduce a method to quantify the degree of reliance on each of the three different sublexical correspondence types using statistical modelling. This technique can be used to test other hypotheses by future studies.

REFERENCES

Andrews, S. (1982). Phonological recoding: Is the regularity effect consistent? Memory & Cognition, 10, 565–575. doi:10.3758/BF03202439
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. doi:10.1037/0096-1523.24.4.1052
Arciuli, J., Monaghan, P., & Seva, N. (2010). Learning to assign lexical stress during reading aloud: Corpus, behavioral, and computational investigations. Journal of Memory and Language, 63, 180–196. doi:10.1016/j.jml.2010.03.005
Brown, G. D., & Deavers, R. P. (1999). Units of analysis in nonword reading: Evidence from children and adults. Journal of Experimental Child Psychology, 73, 208–242. doi:10.1006/jecp.1999.2502
Burani, C., & Arduino, L. S. (2004). Stress regularity or consistency? Reading aloud Italian polysyllables with different stress patterns. Brain and Language, 90, 318–325. doi:10.1016/S0093-934X(03)00444-9
Byrd, R. H., Lu, P., Nocedal, J., & Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. SIAM Journal on Scientific Computing, 16, 1190–1208. doi:10.1137/0911069
Campbell, R., & Besner, D. (1981). This and THAP – Constraints on the pronunciation of new, written words. The Quarterly Journal of Experimental Psychology, 33, 375–396. doi:10.1080/14640748108400799
Cassidy, S., & Harrington, J. (2001). Multi-level annotation in the Emu speech database management system. Speech Communication, 33(1), 61–77. doi:10.1016/S0167-6393(00)00069-8
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. doi:10.1037/0033-295X.108.1.204
Cortese, M. J., & Simpson, G. B. (2000). Regularity effects in word naming: What are they? Memory & Cognition, 28, 1269–1276. doi:10.3758/BF03211827
Cox, F., & Palethorpe, S. (2007). Australian English. Journal of the International Phonetic Association, 37, 341–350. doi:10.1017/S0025100307003192
Duncan, L. G., Castro, S. L., Defior, S., Seymour, P. H., Baillie, S., Leybaert, J., … Francisca Serrano. (2013). Phonological development in relation to native language and literacy: Variations on a theme in six alphabetic orthographies. Cognition, 127, 398–419. doi:10.1016/j.cognition.2013.02.009
Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. Behavior Research Methods, 35(1), 116–124. doi:10.3758/BF03195503
Frost, R., Behme, C., Beveridge, M. E., Bak, T. H., Bowers, J. S., Coltheart, M. (2012). Towards a universal model of reading. Behavioral and Brain Sciences, 35, 263–279. doi:10.1017/S0140525X11001841
Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. Journal of Experimental Psychology: Human Perception and Performance, 5, 674–691. doi:10.1037/0096-1523.5.4.674
Goswami, U. (2002). In the beginning was the rhyme? A reflection on Hulme, Hatcher, Nation, Brown, Adams, and Stuart (2002). Journal of Experimental Child Psychology, 82(1), 47–57. doi:10.1006/jecp.2002.2673
Goswami, U., & Bryant, P. (1990). Phonological skills and learning to read. London: Wiley Online Library.
Goswami, U., Gombert, J. E., & De Barrera, L. F. (1998). Children’s orthographic representations and linguistic transparency: Nonsense word reading in English, French, and Spanish. Applied Psycholinguistics, 19, 19–52. doi:10.1037/S01427164000010560
Goswami, U., Porpodas, C., & Wheelwright, S. (1997). Children’s orthographic representations in English and Greek. European Journal of Psychology of Education, 12, 273–292. doi:10.1007/BF03172876
Goswami, U., Ziegler, J. C., Dalton, L., & Schneider, W. (2003). Nonword reading across orthographies: How flexible is the choice of reading units? Applied
APPENDIX A: GERMAN AND ENGLISH NON-WORDS USED IN EXPERIMENTS 1 AND 2

German
V[C] Reg.
blaf blen (blem in Exp. 1B) blod breg brel brul flom flüb
fryp grät grem grom grul khid klur krel kril krön
pflyp pid plön prod schmün schraf schwüb speg
zwül zwun
V[C][C] Reg.
bamt birt blaff bling boft brals chroll
gärm ginn grabl gunt kall kaxt kerv kluns knell pals
pferrn pfurl purf schern spalf stell sturg zwerp
V[C] Irreg.
bax blex bleg bres flim flis git glef glip krex krin krip pfis
spic stef zwix zwok
V[C][C] Irreg.
bags blags füst gleks kagd kagt kets pagt pard peks poks
schagd stard
English
CS+BR+.
hangst kazz mact phadge phamb phangst plant
sangt slangs slazz stract stramb tamb tazz tradge trazz
zants
CS+BR–.
clett hadl halse kalk kalse phalk phaltz slaltz strald
stralk stralse straltz tald taltz tralse trallt tratt
CS- BR+.
quadge quamb quangst quapse quazz squact squazz
swact swangst swants swazz twadge twangst twants
twazz wact wamb wangs
CS- BR–.
qualk qualse qultz squald squall squalse squaltz
swaltz twald twalk twalse twaltz walz walt whald

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. doi:10.1080/09541440050114570

Ziegler, J. C., Perry, C., Jacobs, A. M., & Braun, M. (2001). Identical words are read differently in different languages. *Psychological Science, 12*, 379–384. doi:10.1111/1467-9280.00370

Ziegler, J. C., Perry, C., Ma-Wyatt, A., Ladner, D., & Schulte-Körne, G. (2003). Developmental dyslexia in different languages: Language-specific or universal? *Journal of Experimental Child Psychology, 86*, 169–193. doi:10.1016/S0022-0965(03)00139-5

Ziegler, J. C., Stone, G. O., & Jacobs, A. M. (1997). What is the pronunciation for -ough and the spelling for/u/? a database for computing feedforward and feedback consistency in English. *Behavior Research Methods, 29*, 600–618. doi:10.3758/BF03210615

Zorzi, M. (2010). The connectionist dual process (CDP) approach to modelling reading aloud. *European Journal of Cognitive Psychology, 22*, 836–860. doi:10.1080/09541440903435621

APPENDIX B: IMPLEMENTING THE FITTING PROCESS IN R

While fitting the models described in the text has a certain flavour of regression to it, there are some important differences. Most critically are the two constraints that we have placed on the parameters: \( \beta_j \in [0,1] \) and \( \sum \beta_j = 1 \). Considerable work has been done to develop and implement estimation methods for models with inequality constraints such as \( \beta_j \in [0,1] \) (Grömping, 2010). However, we know of no such work that has solved the problems presented by the \( \sum \beta_j = 1 \). To address this problem, we turned to the optim function that is part of the base statistical analysis package in R (R Core Team, 2013). optim is a very general optimisation package that allows the user to minimise any specified function, while also placing bounds on the returned values. That is, we can define a function, place upper and lower bounds on the returned weights and optim will efficiently search the allowed parameter space to minimise our function. To satisfy (2), we defined the minimising function to be the residual sum of squares, and restricted the \( \beta \) weights to fall between 0 and 1. This ensures that \( \beta_j \in [0,1] \) is satisfied.

In all of the optimisation analyses, we used the following command in R:

```r
optim (par=runif(3, .2, .8), fn=..., ..., method='L-BFGS-B', control=list(factr=1e5, lower=0, upper=1)).
```

The parameters for optim operate as follows: “par=runif(3, .2, .8)” initialises the \( \beta_j \) s to random values between .2 and .8. “fn=...” specifies the function to be minimised along with any parameters it requires. In our case we used a simple function that calculates the residual sum of squares, “method='L-BFGS-B’” instructs optim to use an optimisation algorithm that allows for upper and lower bounds on the returned values (Byrd, Lu, Nocedal, & Zhu, 1995). “factr=1e5” sets the convergence tolerance, and “lower=0, upper=1” sets the bounds on the returned values.

\[ \sum \beta_j = 1 \text{ Constraint}. \] There is no way to explicitly tell optim to meet the constraint that the \( \beta \) s must sum to 1 (\( \sum \beta_j = 1 \)). One way to ensure that the constraint is met is simply to scale the weights returned by optim using the formula:

\[ \beta_j^* = \beta_j / \sum \beta_j \]

where \( \beta_j^* \) are the new scaled weights, and are guaranteed to sum to 1. However, since this adjustment follows the optimisation process, there is little reason to believe that the resulting \( \beta_j^* \) s would remain an optimal solution to (2). An alternative to simply scaling the \( \beta_j \) s is to make use of the influence of outliers on parameter estimation. For example, according to (2) optim is trying to satisfy the following 180 equations (two per item) simultaneously, by minimising the residual sum of squares (while also meeting the \( \beta_j \in [0,1] \) constraint):
The introduction of a new data point that can only be met by satisfying the constraint that the will put some pressure on optim to select appropriate parameters. For example,

\[ P_1(\text{Short}) = \beta_{\text{gpc}} \times \text{GPC}_{\text{short},1} + \beta_{\text{csc}} \times \text{CSC}_{\text{short},1} + \beta_{\text{brc}} \times \text{BRC}_{\text{short},1} \]

\[ P_1(\text{Long}) = \beta_{\text{gpc}} \times \text{GPC}_{\text{long},1} + \beta_{\text{csc}} \times \text{CSC}_{\text{long},1} + \beta_{\text{brc}} \times \text{BRC}_{\text{long},1} \]

\[ \ldots \]

\[ P_{90}(\text{Short}) = \beta_{\text{gpc}} \times \text{GPC}_{\text{short},90} + \beta_{\text{csc}} \times \text{CSC}_{\text{short},90} + \beta_{\text{brc}} \times \text{BRC}_{\text{short},90} \]

\[ P_{90}(\text{Long}) = \beta_{\text{gpc}} \times \text{GPC}_{\text{long},90} + \beta_{\text{csc}} \times \text{CSC}_{\text{long},90} + \beta_{\text{brc}} \times \text{BRC}_{\text{long},90} \]

Equation (7) is equivalent to creating an artificial data point where all of the dependent and independent variables \([P(\text{Short}), \text{GPC}, \text{CSC}, \text{and BRC}]\) are set to 1. Though (7) provides some pressure to satisfy \(R^b_j = 1\), it is unlikely to have a very large influence since it is only a single equation with roughly equal weight to the other 180. However, dramatically increasing the weight of this data point will exert a much stronger influence on the final parameter selection. For example

\[ 1 = \beta_{\text{gpc}} \times 1 + \beta_{\text{csc}} \times 1 + \beta_{\text{brc}} \times 1 \quad (7) \]

Equation 8 would put enormous pressure on optim to arrive at a set of weights that satisfy \(\Sigma \beta_j = 1\) without putting any further constraints on how the weights are apportioned to the strategies. Though Equation (8) does not guarantee \(\Sigma \beta_j = 1\) precisely, it is sufficiently strong for the present purposes. Other applications may require a larger multiplier.

Finally, because the number of items is not equal across all conditions in our studies, the sums of squares were weighted by item to ensure each condition contributed equally. For example in Experiment 1, items in the \(V[C]\) Irregular and \(V[C][C]\) Irregular conditions received relatively more weight than items in the \(V[C]\) Regular and \(V[C][C]\) Regular conditions. If this isn’t done, there is a tendency for the Regular items to have a stronger influence on the eventual parameters. The weights applied to each item were determined as follows:

\[ \omega_{\text{type}} = \frac{.25}{n_{\text{type}}} \]

where type is one of the four item types (e.g., \(V[C]\) Irregular in Experiment 1), \(\omega_{\text{type}}\) is the weight assigned to items of that type, and \(n_{\text{type}}\) is the total number of items of that type. As this formula implies, each item contributes equally to the influence of its category, but items in smaller categories have more influence than items in larger categories. These weights are then used in the usual weighted sum of squares formula that optim is trying to minimise:

\[ SS_{\text{resid}} = \sum \{Y_i - \bar{Y}_i\}^2 \omega_{\text{type},i} \]