

RUNNING HEAD: LEARNING TO READ AN ARTIFICIAL ORTHOGRAPHY

The Influence of Consistency, Frequency and Semantics on Learning to Read: An
Artificial Orthography Paradigm

J. S. H. Taylor, Kim Plunkett and Kate Nation

University of Oxford

Authors Note

J. S. H. Taylor, Kim Plunkett and Kate Nation, Department of Experimental
Psychology, University of Oxford, Oxford, OX1 3UD, UK.

J. S. H. Taylor is now at the Medical Research Council Cognition and Brain Sciences
Unit, 15 Chaucer Road, Cambridge, CB2 7EF.

This research was supported by an ESRC/MRC studentship awarded to J. S. H. Taylor.

We thank the reviewers for their helpful comments on earlier versions of this article.

Correspondence concerning this article can be addressed to J. S. H. Taylor

(joanne.taylor@mrc-cbu.cam.ac.uk) or Kate Nation (kate.nation@psy.ox.ac.uk).

Abstract

Two experiments explored learning, generalization and the influence of semantics on orthographic processing in an artificial language. In Experiment 1, 16 adults learned to read 36 novel words written in novel characters. Post-training, participants discriminated trained from untrained items and generalized to novel items, demonstrating extraction of individual character-sounds. Frequency and consistency effects in learning and generalization showed that participants were sensitive to the statistics of their learning environment. In Experiment 2, 32 participants were pre-exposed to the sounds of all items (*lexical phonology*) and to novel definitions for half of these items (*semantics*). Pre-exposure to either lexical phonology or semantics boosted the early stages of orthographic learning relative to Experiment 1. By the end of training, facilitation was restricted to the semantic condition and to items containing low frequency inconsistent vowels. Pre-exposure reduced generalization suggesting that enhanced item-specific learning was achieved at the expense of character-sound abstraction. Our novel paradigm provides a new tool to explore orthographic learning. While our findings support the idea that semantic knowledge supports word reading processes, they also suggest that item-specific phonological knowledge is important in the early stages of learning to read.

Keywords: artificial orthography, reading, statistical learning, semantics, lexical phonology

The Influence of Consistency, Frequency and Semantics on Learning to Read: An Artificial Orthography Paradigm

One of the most impressive aspects of the human language faculty is our ability to cope with both rules and exceptions to those rules. This is particularly well illustrated by English orthography. A skilled reader of English is able to read words that follow typical spelling-sound patterns, i.e., regular or consistent words, and words that do not, i.e., exception or inconsistent words. They can also assign a pronunciation to novel orthographic forms. How we learn to balance the twin demands of a system that promotes creativity and generalization while at the same time allowing exceptions has been considered extensively in the psycholinguistic literature (Fodor & Pylyshyn, 1988; Glushko, 1979; Humphreys & Evett, 1985; Pinker, 1991) and has been the focus of a number of modelling initiatives using a variety of different architectures (Coltheart, Curtis, Atkins, & Haller, 1993; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Harm & Seidenberg, 1999;2004; Jacobs & Grainger, 1994; Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg & McClelland, 1989; Zorzi, Houghton, & Butterworth, 1998). At the same time, good progress has been made towards understanding the beginnings of reading development in children, but relatively little research has examined the learning process directly. In this paper, we describe two experiments that exposed adults to novel words written in an artificial orthography. These allowed us to address four main issues. First, can learners extract sub-word regularities (akin to grapheme-phoneme correspondences) from exposure to whole-word orthographic forms (and their corresponding pronunciations) without explicit instruction? Second, are learners

sensitive to lexical properties such as spelling-sound consistency and frequency, present in the language environment? Third, can learners generalize to novel forms, and is generalization influenced by the frequency and consistency characteristics of the training set? Finally, how does the introduction of meaning to the language affect learning and generalization?

Our study was inspired by the growing number of studies using artificial languages to explore issues in spoken language processing. For example, Magnuson, Tanenhaus, Aslin and Dahan (2003) exposed adults to new phonological forms that varied in frequency and phonological overlap. This allowed them to chart the emergence of neighbourhood effects as new forms competed with those already established. Wonnacott, Newport and Tanenhaus (2008) used similar techniques to investigate grammatical learning. They exposed adults to a new language in which verbs occurred probabilistically in one of two constructions. Post-tests demonstrated that adults had learned and abstracted the statistical regularities governing verb use in the language and could use them productively and in on-line comprehension.

Why should we wish to use an artificial language to study word reading? Primarily, it allows precise control over the input statistics to which learners are exposed. This avoids the need to rely on lexical databases to extract variables such as frequency and consistency which can only ever be a proxy for an individual's experience with the language. It also reduces concerns about other non-controlled factors influencing performance. As noted by Magnuson (2008), this point is nicely illustrated by the apparent disappearance of a consistency x imageability interaction in reading aloud once age-of-acquisition is controlled (Monaghan & Ellis, 2002). In addition it combats the problem of using the restricted word lists that result from

attempting to control for all of these factors (Forster, 2000). While the natural correlation between psycholinguistic variables may in fact reflect what is optimal, it is important to understand how each operates before addressing such higher level questions. The degree of control provided by an artificial language therefore makes it an ideal tool to question and extend the existing literature on how lexical statistics influence learning and generalization.

A surprisingly small number of studies have used such methodologies to investigate issues concerned with single word reading. McCandliss, Posner and Givón (1997) taught adults novel words which were assigned the same meanings as existing English words and spelled using the Roman alphabet. Although their language was well-suited to explore the issues of lexicalisation that they set out to investigate, it was not artificial either in terms of orthography or meaning. A different approach is exemplified by Bowers, Davis and Hanley (2005). They taught new novel forms, such as BANARA, that were related in form to so-called hermit words: words with no orthographic neighbours (e.g. BANANA). After training, they examined whether learning of new forms competed with the processing of hermits. Once again, this was not an artificial orthography, but based on close neighbours of existing words written in standard form.

A small cluster of studies emerged in the 60s and 70s using novel symbols (for a review, see Knafle & Legenza, 1978) but these tended to focus on single symbol-sound correspondences, rather than symbol sequences forming words (see also Byrne, 1984; Byrne & Carroll, 1989). This was also the case in some more recent neuroimaging studies examining changes in brain activity pre- and post-training on previously unknown orthographic symbols (Callan, Callan, & Masaki, 2005;

Hashimoto & Sakai, 2004). Two experiments that have used word-level forms presented in an artificial orthography are described by Bitan and Karni (2003; 2004). They asked whether adults could extract sub-word spelling-sound correspondences from novel words written in a novel orthography and concluded that they could not. Although participants were able to learn the training sets, they were unable to generalize their knowledge to novel forms written in the same novel orthography without explicit teaching in symbol-sound correspondences.

On first sight, these findings cast doubt on whether an artificial language paradigm has any utility for investigating orthographic learning. Importantly however, there are several problems with Bitan and Karni's methodology which may undermine the utility of their approach. First, unlike most alphabetic scripts in which phonemes are represented by a cohesive symbol, phonemes in these studies were represented by two or three separate symbols. For example, the sound /p/ was written as /[^]*, /b/ as */ and /n/ as [^]/[^], meaning that the word /pɒn/ was written as /[^]**/[^]/[^]. This complexity was used to minimise the impact of existing alphabetic knowledge. However, it is likely to have actively hindered the extraction of sub-word symbol-sound knowledge. Second, in natural languages, children are soon faced with an ever increasing set of words, whereas, Bitan and Karni taught adults only 6-12 words. Coupled with the intractable nature of the sub-word symbol-sound relationships, it is not surprising that learners adopted a whole-word rote learning strategy. Third, learning involved participants deciding whether their decoding attempts matched an English translation, and making same-different judgements about pairs of stimuli. This is very different to the corrected pronunciation attempts

that typically characterise how children learn to read. Together, these factors may well have promoted very different learning to that seen when people acquire a natural alphabetic script. Thus, whether learners can extract symbol-sound correspondences from an artificial orthography and use them to support generalization to new forms remains an open question.

Models of reading aloud also provide valuable information as to how we might cope with rules and exceptions and generalize to novel forms. In the Dual-Route Cascaded model (DRC) (Coltheart et al., 2001) there are two routes to reading. The non-lexical route stores rules for converting graphemes into phonemes and is essential for reading novel words. The lexical route stores whole-word orthographic forms and their pronunciations and is essential for reading irregular words which would be pronounced incorrectly using grapheme-phoneme correspondence rules (GPCs). Both routes can pronounce regular words correctly.

GPCs were selected on the basis of frequency in the language, with the most common pronunciation of a grapheme being considered its phoneme correspondence. Each GPC maps a grapheme to a single phoneme and is largely insensitive to context, i.e. what the other letters in a word are. For example, GPCs do not capture the fact that, although the most common pronunciation of l is /l/ as in FISH, l is pronounced /aɪ/ in the overwhelming majority of words that end in ND, e.g., FIND, MIND and KIND. Instead these words are classed as irregular and they are stored as whole-word forms by the lexical route. This seems somewhat counterintuitive and, in fact, Treiman, Kessler and Bick (2003) demonstrated that we are sensitive to context when reading novel words. For example, they found that the

nonword CHIND was usually pronounced similarly to FIND, MIND and KIND rather than CHINK.

The DRC is explicitly non-developmental, with Coltheart and colleagues avoiding learning issues on the grounds that "... unless the learning procedure itself is known to be psychologically real, it may not be able to learn what people learn..." (Coltheart et al., 2001, p. 216). While this is certainly the case, thinking about how systems learn is important. Possessing context insensitive GPCs, selected on the basis of frequency in the language, served the DRC well, but they may not capture what children learn about spelling-sound mappings. When children first come to the task of learning to read they do not know which words are regular and which irregular. Research has not directly examined how this knowledge develops; presumably it emerges over time and is extracted from experience with reading words. However, this seems at odds with Bitan and Karni's (2003; 2004) findings that adults are unable to do this in an artificial orthography. Clearly, questions remain regarding whether learners can extract sub-word regularities without explicit teaching.

In contrast to the DRC, Parallel Distributed Processing models, for example the triangle model (Harm & Seidenberg, 2004; Plaut et al., 1996), are explicitly developmental. The triangle model comprises sets of units coding for phonological, orthographic and semantic information. It learns to read by being presented with the orthographic form of a word, producing a pronunciation attempt and receiving the correct pronunciation as feedback. This feedback then modifies the strength of the connections between units, thereby increasing the probability that future pronunciation attempts will be correct. Importantly, the model has no inbuilt GPCs

and develops context sensitivity, pronouncing nonwords such as CHIND similarly to FIND, MIND and KIND, rather than CHINK. The model does not dedicate different units to the representation of different word types (i.e., words vs. nonwords, or regular words vs. irregular words) and instead, effects of lexicality, frequency and context-sensitivity emerge as a consequence of statistical learning from exposure to the language. This questions the necessity of the DRCs distinction between non-lexical and lexical reading processes and whether human learners use GPCs.

Although the psychological validity of the learning mechanisms is debated (Coltheart, 2005; Harm & Seidenberg, 1999), Plaut et al. (1996) demonstrated the importance of considering learning when studying the structure of the reading system. However, as Coltheart et al. (2001) note, such modelling simulations do not tell us how humans learn, they only make predictions. Thus, in Experiment 1 we investigated whether learners are sensitive to statistical patterns embedded within a novel orthography. This is a first step towards assessing the principle that sub-word regularities can be extracted following exposure to language, with sensitivity to frequency and consistency “falling out” of this exposure.

Another factor that has been implicated in word reading is meaning. Coltheart et al. (2001) suggested that the lexical route of the DRC further subdivides into a semantic and a non-semantic pathway but made no specific predictions as to how this might operate. In contrast, both empirical studies (e.g., McKay, Castles, Davis, & Savage, 2007; Patterson et al., 2006; Pexman, Hargreaves, Siakaluk, Bodner, & Pope, 2008; Strain, Patterson, & Seidenberg, 1995; Woollams, 2005; Woollams, Lambon-Ralph, Plaut, & Patterson, 2007) and the triangle model (Harm & Seidenberg, 2004; Plaut et al., 1996) posit a role for semantics in word reading that is

particularly important for low frequency words with inconsistent spelling-sound mappings. However, as we discuss in the Introduction to Experiment 2, evidence is rather mixed particularly when developmental research is considered and so in Experiment 2, we use the control provided by the artificial orthography paradigm to investigate the role of semantics in *learning* to read words.

Experiment 1

Experiment 1 considered whether adults are sensitive to statistical regularities present in a novel orthography following a relatively short exposure to the written and spoken forms of whole words. The experiment focused specifically on the influence of spelling-sound consistency and frequency on learning and generalization. In the existing literature the term consistency refers to the predictability of a word's spelling-sound mapping. It is defined on the basis of the central vowel plus following consonants (orthographic body) and is therefore sensitive to context, unlike GPCs. Glushko (1979) described consistent words as those containing an orthographic body with only one possible pronunciation in English (e.g., RINK) and inconsistent words as those containing bodies that have more than one possible pronunciation (e.g., BOOK, SPOOK). Jared, McRae and Seidenberg (1990) developed this further, quantifying consistency as the number of friends vs. enemies a word possessed. Using this measure, BOOK is a highly consistent word as it has many friends, e.g., TOOK, LOOK, COOK, ROOK, and only one enemy, SPOOK, a highly inconsistent word. This better reflects the graded nature of spelling-sound relationships in English orthography.

Consistent words are responded to faster than inconsistent words in both naming (Cortese & Simpson, 2000; Jared, 1997;2002) and lexical decision (Lacruz & Folk, 2004; Stone, Vanhoy, & VanOrden, 1997). In naming, this effect is modulated by word frequency: high frequency words are less detrimentally influenced by spelling-sound inconsistency than low frequency words, (Andrews, 1982; Seidenberg, 1985; Seidenberg, Waters, Barnes, & Tanenhaus, 1984; Taraban & McClelland, 1987; Waters & Seidenberg, 1985). The interaction between frequency and consistency is somewhat debatable in lexical decision (Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Lacruz & Folk, 2004). Finally, as described earlier, Treiman et al. (2003) demonstrated that adults take consistency into account when generalizing to novel words.

We used an artificial orthography learning paradigm to explore the influence of spelling-sound consistency and frequency on reading for two reasons. First, it enabled complete control over exposure to the statistics of the language permitting greater precision of and confidence in results than those from research on natural languages. Second, it provided an environment to examine learning, allowing us to assess whether sub-word regularities can be extracted purely through exposure to whole-word pronunciations.

Participants

Sixteen adults (6 males, 10 females) took part. They were all university students and their mean age was 20 years, $SD = 2.34$. All participants had English as a first language and had normal, or corrected to normal, hearing and vision.

Materials and Procedure

All participants completed the same three phase procedure: first, an exposure and learning phase, second, an old-new decision task that involved discriminating trained from untrained items, and third, a generalization task that asked them to read aloud a set of untrained items.

Exposure and learning phase. Two sets of 36 training items were constructed. Half the participants were exposed to one set and half to the other in order to minimise the impact of any idiosyncratic properties of the items. All training items were monosyllabic consonant-vowel-consonant novel words. The items in each set were constructed from 12 consonant and 6 vowel phonemes. The consonants were the same in both sets (/b/, /d/, /f/, /g/, /k/, /m/, /n/, /p/, /s/, /t/, /v/, /z/) but the vowels differed. Set one contained the vowels /ɛ/, /i/, /aɪ/, /ɒ/, /əʊ/, /u/, and set two /ɪ/, /i/, /aɪ/, /əʊ/, /ʌ/, /u/. The written forms of both sets of training items were constructed from the same 16 characters. However, to minimise any effects of particular phonemes being easier to map to particular characters, two different randomly assigned character-phoneme mappings were used and half the participants who were exposed to each training set experienced one set of character-phoneme mappings and half the other, see Appendix A.

In both training sets consonant phonemes were represented by a single character. However, vowel character-phoneme mappings varied in consistency. Two vowel characters in each set were *consistent* and pronounced in the same way in all items. The other two vowel characters were inconsistent and were pronounced one way when preceded by a particular consonant character (*inconsistent-conditioned* pronunciation) and in a different way when preceded by any of the other consonant

characters (*inconsistent-unconditioned* pronunciation). In English orthography, vowel pronunciations are more often affected by final than initial consonants (Treiman, Mullennix, Bijeljacabic, & Richmond-Welty, 1995) whereas in our novel orthography, the reverse was true. This meant that learning about the conditional relationships between consonant and vowel pronunciations could not be based on existing knowledge of English.

We also explored whether the influence of vowel consistency on learning varied as a function of frequency. One of the *consistent* characters was *high frequency*, occurring in eight items, and one *low frequency*, occurring in four items. One inconsistent character was pronounced in the *inconsistent-conditioned* way in eight items and the *inconsistent-unconditioned* way in four items, while the other inconsistent character was pronounced in the *inconsistent-conditioned* way in four items and the *inconsistent-unconditioned* way in eight items.

It should be highlighted that whereas we explored the influence of vowel type frequency: consistent and inconsistent vowel types occurred in many or few words, the existing literature has examined word token frequency effects by contrasting consistent and inconsistent words that occur often or rarely in the language. Our frequency manipulation allowed us to examine how the impact of consistency on orthographic learning was modulated by how often a character-sound mapping was experienced. However, there are possible differences between the influences of word and vowel frequency manipulations which will be further explored in the General Discussion.

Except for the two consonants that formed the onset for inconsistent-conditioned items which necessarily occurred as often as the vowel phonemes they

affected (i.e., 4 or 8 times), consonants occurred approximately the same number of times in onset position across training items. Similarly, all consonants occurred approximately the same number of times in coda position. Full details of the consistency x frequency manipulation for one of the training sets are given in Appendix B, along with character-phoneme mappings; their pronunciations are provided in Appendix C, Table C1. Participants viewed written stimuli on a monitor and heard spoken stimuli (recorded by a female speaker and digitised at a sampling rate of 44Hz) through headphones. They wore a microphone and were recorded throughout the experiment.

In the exposure phase participants viewed each training item, listened to its pronunciation and repeated it once. The item remained on the screen until the participant had repeated it, and the experimenter had recorded whether their pronunciation was correct. Presentation was randomised. Participants then moved into the training phase in which they viewed each training item and attempted to read it. Response times were not restricted. Once participants had provided a response the correct pronunciation was given as feedback and the experimenter recorded whether their attempt was correct. All items were presented in a randomised order. If less than 70% of items were read correctly, all items were attempted again. This criterion was chosen because pilot studies indicated that at this level participants were able to competently complete post-tests.

Old-new decision. After training we assessed participants' ability to discriminate the orthographic forms of 12 trained items (targets) from 12 untrained but similar items (distractors). The untrained distractors were the same as those used to assess generalization, see below. An item was presented on the screen and

participants pressed “z” if they judged it to be one they had learned and “m” if they judged it to be one they had not learned. Response times were not restricted but participants were instructed to respond as quickly as possible. They were also told that the task would be very hard because all items would look extremely similar. Accuracy and reaction times (RTs) were recorded.

Generalization. A set of 12 generalization items was created for each training set. Four items in each set contained consistent vowel characters and eight contained inconsistent vowel characters. Of the eight items containing inconsistent characters, four were preceded by a consonant character that should cause the vowel to take the inconsistent-conditioned pronunciation and four by a consonant character that should cause the vowel to take the inconsistent-unconditioned pronunciation. Half the test items contained consistent, inconsistent-conditioned and inconsistent-unconditioned vowels that were high frequency during training and half contained vowels that were low frequency during training. This enabled us to assess how the frequency and consistency of vowels during training influenced generalization. Appendix C, Table C2 gives an example of one of the generalization sets.

Participants attempted to read aloud each of the generalization items. Response times were not restricted and no feedback was given. Items were scored correct if consonants were pronounced correctly and if vowels were pronounced according to their consonant context, i.e., given the inconsistent-conditioned or inconsistent-unconditioned pronunciation when appropriate. Training, old-new decision and generalization were all completed in a single session lasting between 30 and 45 minutes.

Results

Training

Participants varied in the number of training blocks they needed to achieve $\geq 70\%$ accuracy ($M = 3.63$ blocks, $SD = 1.45$, $max = 6$). The large standard deviations seen in Figure 1 show that participants varied greatly in reading accuracy during training, particularly in earlier blocks. There was a substantial negative correlation between accuracy in Block 1 and the number of blocks to achieve at least 70% accuracy, $r = -.72$, $p < .01$, i.e., those who performed best at the beginning learned fastest.

INSERT FIGURE 1 ABOUT HERE

The effects of vowel frequency and consistency were assessed by conducting subjects (F_p) and items (F_i) analyses of variance (ANOVAs) on the proportion of items read correctly in the block in which at least 70% accuracy was achieved. This constituted different blocks for different participants. It should also be noted that while each participant learned a total of 36 items, across participants there were 72 items (two sets of 36 items). These results are summarised in Figure 2.

Accuracy was significantly higher for items containing high frequency vowels than for those containing low frequency vowels, $F_p(1, 15) = 23.00$, $p < .001$, $F_i(1, 66) = 35.47$, $p < .001$. It was also higher for items containing consistent vowels, than for those containing inconsistent-conditioned vowels and these in turn outperformed items containing inconsistent-unconditioned vowels, $F_p(2, 30) = 12.65$, $p < .001$, $F_i(2, 66) = 19.14$, $p < .001$. The interaction between vowel frequency and consistency was significant, $F_p(2, 30) = 3.48$, $p < .05$, $F_i(2, 66) = 6.72$, $p < .001$. Pairwise comparisons demonstrated that the facilitative effect of frequency was significant in items

containing inconsistent vowels but not in items containing consistent vowels.

Similarly, the consistency effect was significant in items containing low frequency vowels but not in items containing high frequency vowels.

INSERT FIGURE 2 ABOUT HERE

Old-new decision

All data points were included for analyses of discrimination accuracy. For latency analyses, only correct responses were considered and trials with RTs more than two standard deviations away from the participant's mean were excluded (2.9 %). One sample t-tests demonstrated that accuracy was significantly above chance on both trained, $t_p(15) = 12.96, p < .001$, and untrained items, $t_p(15) = 4.56, p < .001$. However, the proportion of false alarms (trials in which untrained items were incorrectly judged to be trained items) was relatively high, $M = .37 (SD = .11)$, and accuracy was significantly better for trained, $M = .81 (.09)$, than untrained items, $M = .63 (.11), t_p(15) = 4.51, p < .001$. RTs (ms) were also faster for trained, $M = 3775 (989)$, than untrained items, $M = 4743 (1369), t_p(15) = 5.08, p < .001$.

Two ANOVAs assessed the influence of vowel frequency and consistency on old-new decision, one with proportion of trained items correct as the dependent variable and one using mean RTs to correct trained items. One participant had no correct responses to items containing inconsistent-conditioned low frequency vowels and one had no correct responses to items containing inconsistent-unconditioned high frequency vowels. This led to two missing data points in the RT dataset; these were replaced with the mean RT for that item type. There were no main effects or interactions in the accuracy analysis (Frequency, $F_s < 1$, Consistency, $F_s < 1$, Frequency x Consistency, $F_p(2, 30) = 2.36, p > .10, F_i < 1$). However, RTs were

faster to items containing high frequency vowels, $M = 3420\text{ms}$, than to items containing low frequency vowels, $M = 4193\text{ms}$, $F_p(1, 15) = 6.68$, $p < .05$, $F_i(1, 18) = 9.76$, $p < .01$. Neither the main effect of consistency, $F_s < 1$, nor the consistency x frequency interaction, $F_s < 1$, was significant in the RT analysis.

Generalization

The mean proportion of untrained test items read correctly was .71 (.13). An ANOVA assessed the impact of vowel frequency and consistency during training on the proportion of generalization items read correctly. Results are summarised in Figure 3. Accuracy was higher for items containing high frequency vowels, $F_p(1, 15) = 8.04$, $p = .01$, $F_i(1, 18) = 12.14$, $p < .01$. Accuracy was also higher for items containing consistent vowels relative to those containing inconsistent-unconditioned vowels, $F_p(2, 30) = 5.29$, $p = .01$, $F_i(2, 18) = 5.97$, $p = .01$. Accuracy for items containing inconsistent-conditioned vowels did not differ from items containing either of the other two vowel types. The interaction between vowel frequency and consistency was not significant, $F_p < 1$, $F_i(2, 18) = 1.07$, *ns*, but Figure 3 shows a clear trend towards an interaction mirroring that seen in during training.

INSERT FIGURE 3 ABOUT HERE

Discussion

The results of Experiment 1 show very clearly that adults extracted the sounds of individual characters following exposure to whole-word forms, and that they were able to use this knowledge to read novel items successfully. In addition, the frequency and consistency of character-sound mappings influenced learning and generalization, showing that participants were sensitive to the statistical properties

of their learning environment. Specifically, trained items were read more accurately when they contained consistent and/or high frequency vowels. In addition, the effects of consistency and frequency interacted; consistency only affected items containing low frequency vowels, and the advantage of items containing high over low frequency vowels was only present when vowels were inconsistent. These effects transferred to generalization: accuracy was higher for generalization items containing vowels that were consistent and/or high frequency during training and there was a trend for an interaction between vowel frequency and consistency which mirrored that found in trained items. These findings support the idea that learners can extract context sensitive sub-word regularities from exposure to whole-word forms. Furthermore, the effects of consistency and vowel type frequency resembled consistency and word token frequency effects observed in natural languages. This validates the use of an artificial orthography paradigm for further investigating the factors affecting learning to read.

Two additional findings are of interest. First, reading accuracy was higher for items containing conditioned inconsistent vowels than for items containing unconditioned inconsistent vowels, when vowels were also low in frequency. This is in line with Kessler and Treiman's (2001) finding that the predictability of vowel pronunciations in English words is increased by taking the consonant context into account. Our finding furthers Kessler and Treiman's observations by demonstrating that vowel predictability facilitates both learning and generalization, and is sensitive to frequency. Second, at the end of training, participants were able to discriminate trained from untrained items, and they were faster and more accurate at dealing with the more familiar forms. In typical lexical decision tasks, on which the old-new

task was based, latencies are also sensitive to item frequency and consistency (Lacruz & Folk, 2004; Stone et al., 1997). Although old-new decisions were faster to items containing high frequency vowels, they were not affected by item consistency. In addition, false positive rates were fairly high and RTs were extremely long (in the order of seconds rather than milliseconds). This suggests that there are some differences between typical lexical decision and old-new decision in the artificial orthography. This issue is considered further in the General Discussion.

One clear difference between real words and the items in our artificial orthography is that our items had no meaning. Potentially, an absence of semantics in the language may have promoted atypical learning. For example, it may have enhanced extraction of sub-word character-sound relationships at the expense of whole item learning. This issue was addressed in Experiment 2.

Experiment 2

Supporting a semantic effect on reading aloud, Strain, Patterson and Seidenberg (1995) reported an imageability x frequency x consistency interaction: adults were slower and more error prone when reading low frequency exception words of low imageability relative to low frequency exception words of high imageability. This finding was replicated by Frost et al. (2005). Imageability also influences lexical decision speed (Balota et al., 2004; Cortese & Khanna, 2007), although no interaction between semantics and frequency or consistency is typically reported in this task. Corroborating evidence is provided by the finding that patients with semantic dementia have difficulties in reading aloud, especially when words are

low in frequency and inconsistent (McKay et al., 2007; Patterson et al., 2006; Woollams et al., 2007).

The DRC model proposes that known words can be read either by a semantic or a non-semantic lexical route. However, as the semantic lexical route has not been implemented, the DRC makes no explicit predictions as to how a semantic contribution to reading aloud operates. In contrast, the triangle model makes specific predictions concerning the role for semantics. In this model, there are direct connections between orthography and phonology, but also indirect connections that enable semantics to influence the computation of phonology from orthography. Simulations by Plaut et al. (1996) demonstrated that the indirect (semantic) pathway was utilised more when reading low frequency inconsistent words. As such words have atypical and uncommon mappings between orthography and phonology, they place considerable strain on the direct route. A boost from semantics therefore increases the system's efficiency when reading these words.

Although the view that semantics plays a role in single word reading enjoys some support from empirical studies, there are several reasons why this conclusion is not unequivocal. First, studies are plagued by the problems outlined earlier concerning natural correlations between lexical variables. To reiterate a specific example that is directly relevant to the present discussion, Monaghan and Ellis (2002) found that the effect of imageability on reading aloud (Strain et al., 1995) disappeared when age of acquisition (AoA) was taken into account. This suggests that the facilitation seen for high imageability words may in fact be driven by a reading aloud advantage for words learned early in life (Gilhooly & Watson, 1981). Second, the strength of the evidence from neuropsychology is diminished by

observations of intact inconsistent word reading in patients with semantic dementia (Blazely, Coltheart, & Casey, 2005; Schwartz, Marin, & Saffran, 1979). Taken together, these factors call into question the extent to which inconsistent word reading is dependent on semantics.

Finally, perhaps the most substantial unresolved issue in this literature concerns the difficulty of distinguishing between the effect of knowledge of word meanings (semantics) and the effect of knowledge of the sounds of words (lexical phonology). Several authors have argued that AoA effects arise from differences in the strength of phonological representations (Brown & Watson, 1987; Morrison & Ellis, 1995;2000; Morrison, Hirsh, Chappell, & Ellis, 2002), although see Bonin, Barry, Meot and Chalard (2004) and Cortese and Khanna (2007) for alternative views. If AoA in part exerts its effects by influencing phonological representations, the finding that AoA drives imageability effects suggests that familiarity with the sound of a word might be crucial for reading aloud, rather than familiarity with its meaning. Against this, subsequent work has shown that imageability effects are reduced in tasks designed to minimise semantic involvement (Woollams, 2005), consistent with there being a semantic locus to the imageability effect. Clearly however, it is difficult to separate the influence of familiarity with lexical phonology vs. familiarity with semantics as these two variables are very highly correlated in natural language.

To address this issue, McKay, Davis, Savage and Castles (2008, Experiment 2) familiarised adults with the phonological form of 20 novel words and with meanings for half of these words. For example, they were taught that /nɪlt/ meant “A cape worn by a bullfighter”. They then learned to read the novel words, half of which had consistent spelling-sound correspondences and half inconsistent. Semantic pre-

training boosted the initial stages of orthographic learning and effects were specific to inconsistent items. Participants also made fewer errors and were faster to respond to semantic items in speeded naming, and were more accurate at recognising them in an old-new decision task. McKay et al. (2008) concluded that semantics plays a small but significant role in reading aloud inconsistent words, over and above any effects of lexical phonology.

One potential problem is that McKay et al. (2008) did not include a baseline no pre-exposure condition. This means that pre-exposure to lexical phonology may also have aided orthographic learning, albeit to a lesser extent than semantic pre-exposure. A further issue is that McKay et al. used English orthography and participants were highly proficient from the beginning of training. This is important because developmental research suggests semantic effects may differ at lower levels of proficiency. For example, McKague, Pratt and Johnston (2001) found that 6- to 7-year-old children learned to read novel words more successfully if they had been pre-exposed to their phonological forms and that semantic pre-exposure provided no additional benefit. In their experiment, all items had consistent spelling-sound mappings. In contrast, Nation and Cocksey (2009) found a predictive relationship between phonological familiarity and reading aloud that only held for irregular words. 7-year-old children were better able to read irregular words that they had recognised in an auditory lexical decision task, but the ability to provide a definition for these items offered no additional benefit. Overall these developmental studies suggest that knowledge of word sounds rather than word meanings may benefit the early stages of learning to read. They also leave open the possibility that the benefit conferred by phonological familiarity may not be specific to inconsistent words.

Experiment 2 investigated the role of semantic vs. phonological familiarity in learning to read using the artificial orthography paradigm introduced in Experiment 1. The paradigm is ideal for examining semantic effects on reading aloud because the frequency and consistency of character-sound mappings have such clear effects on performance. It also enabled complete control over familiarity with semantics versus lexical phonology, and results could be compared to those from Experiment 1, a no pre-exposure condition. Furthermore, since participants are unfamiliar with the orthography prior to the experiment, we were able to examine semantic effects at both low and high levels of reading proficiency. A different set of adults completed the same experimental procedure as in Experiment 1. However, prior to this, participants were pre-exposed to the sounds of all the items (lexical phonology) and to novel definitions for half the items (semantics).

Experiments with adults suggest that by the end of training, semantic familiarity should improve reading aloud and old-new decision, relative to both phonological and no familiarity (McKay et al., 2008). In reading aloud these effects should be specific to items with low frequency inconsistent character-sound mappings. In contrast, developmental evidence suggests that phonological familiarity should benefit the early stages of orthographic learning (potentially irrespective of frequency/consistency manipulations) and that semantic knowledge may not provide additional benefit.

Participants

Thirty-two native English speaking university students (9 males, 23 females) took part in Experiment 2. Their mean age was 22.06 years, $SD = 3.24$. Participants had normal hearing and vision and had not participated in Experiment 1.

Materials and Procedure

Participants completed the same procedure as in Experiment 1. Prior to this, they completed a pre-exposure phase that involved learning the sounds for all the items (lexical phonology) and a definition for half of the items (lexical phonology + semantics). Definitions were adapted from Oxford English Dictionary entries of extremely low-frequency concrete English words. Seventeen of these were taken from McKay et al. (2008), and one “an assistant to a magician or scholar” was added. For each participant, 18 items were assigned to the lexical phonology condition and 18 to the semantic condition. Items containing high and low frequency, consistent, inconsistent-conditioned and inconsistent-unconditioned vowels were evenly distributed between these two conditions. Sixteen participants received one half of the items in the lexical phonology condition and the other half in the semantic condition, and the remaining 16 participants received the reverse assignment.

Pre-exposure phase. Participants were told that they would be learning how to say some new words and that for half the words they would also be learning their meanings. They were specifically asked to try and remember the meanings of the words and were told that they would be tested on this. On each trial, a fixation cross appeared on the screen and participants listened to and repeated one of the training items. In the lexical phonology condition this constituted one trial and the next item was then presented. In the semantic condition, a written definition then appeared on screen and participants were instructed to read the definition either silently or out loud and to press spacebar to move onto the next trial. Participants experienced six semantic trials, followed by six lexical phonology trials; this process was then repeated twice so that each of the 36 training items had been presented. This

constituted one pre-exposure block. The order of items within the lexical phonology and semantic items was randomised within a block.

After three pre-exposure blocks were completed, participants heard each of the 18 semantic items and were asked to say its definition out loud. After they had provided a response, the correct definition appeared on screen and participants pressed spacebar to move onto the next recall trial. Item presentation was randomised. To ensure that the 18 lexical phonology items were heard as many times as the semantic items during the pre-exposure phase, participants then completed a series of six serial recall tasks. These involved listening to three lexical phonology items and repeating them back in the correct order. Each of the 18 lexical phonology items was presented once and item presentation was randomised.

Following these semantic and lexical phonology recall tasks, three more pre-exposure blocks were completed, and the semantic and lexical phonology recall tasks were again administered. This completed the pre-exposure phase.

Orthography training and testing. Following pre-exposure, training and testing on the novel orthography continued as in Experiment 1 but with two differences. First, all 36 training items were included in the old-new decision task to ensure that power was high enough to examine the impact of the within subject semantic vs. lexical phonology manipulation. 24 distractor items were also included. Twelve of these distractors were the untrained test items described in Experiment 1, and 12 were additional items constructed in the same way. Second, after old-new decision and before generalization, we checked participants' memory for the definitions. Participants heard each of the semantic items, presented in a random order, and were asked to recall its definition out loud. Unlike during pre-exposure,

no feedback was given. The whole procedure was completed in a single session lasting around 60 minutes.

Results

Semantic learning

A score of 1 was given when the key features of a definition were recalled (e.g., “a wooden case used for storing cannonballs” recalled as “a box for cannonballs”). A score of 0 was given when the wrong or no definition was recalled, or when only minimal features were remembered (e.g., “the fold of skin hanging from the neck of cattle” recalled as “something to do with a cow”). The mean proportion of definitions correctly recalled was exactly the same at the end of pre-exposure and following the old-new decision task, $M = .93$ (.25) demonstrating that the meanings of the semantic items had been learned and remembered.

Beginning of training

Lexical phonology vs. definitions. An ANOVA examined the effect of pre-exposure (lexical phonology vs. definitions) on reading accuracy in the first block of training as a function of vowel frequency and consistency. The results of this analysis are summarised in Figure 4 and further broken down in Table 1. Accuracy was higher for items containing high frequency vowels than for those containing low frequency vowels, $F_p(1, 31) = 23.97, p < .001, F_i(1, 66) = 16.97, p < .001$. The main effect of consistency was also significant, $F_p(2, 62) = 3.81, p < .05, F_i(2, 66) = 3.83, p < .05$. Accuracy was higher for inconsistent-conditioned than inconsistent-unconditioned items and accuracy for consistent items fell somewhere in between. There was no interaction between vowel frequency and consistency at this early stage of training, $F_p(2, 62) = 1.64, ns, F_i(2, 66) = 1.15, ns$. The main effect of pre-exposure was not

significant, $F_p(1, 31) = 1.73$, *ns*, $F_i(1, 66) = 1.01$, *ns*. Pre-exposure did not interact with vowel frequency, $F_s < 1$, or consistency, $F_s < 1$, and the three-way interaction between these variables was also non-significant, $F_s < 1$.

INSERT FIGURE 4 AND TABLE 1 ABOUT HERE

Definitions/lexical phonology vs. no pre-exposure. We next compared performance in the definitions and lexical phonology conditions to the no pre-exposure baseline provided by Experiment 1 in two separate analyses of variance. Relevant data are again contained in Figure 4 and Table 1. It was not possible to combine these three conditions in one analysis because the semantic vs. lexical phonology manipulation was within subject, whereas the comparison with Experiment 1 was between subjects. Note that 16 individuals participated in Experiment 1 whereas 32 individuals participated in Experiment 2. However, for each participant in Experiment 2, half the items contributed to the definitions condition and half to the lexical phonology condition. Therefore, the overall power was similar across the no pre-exposure and definitions/lexical phonology conditions. This design was also used to examine the end of training data reported in the next section.

The first analysis demonstrated that accuracy was higher following definitions pre-exposure than in the no pre-exposure baseline provided by Experiment 1, $F_p(1, 46) = 18.07$, $p < .001$, $F_i(1, 66) = 83.20$, $p < .001$. The effect of pre-exposure did not interact with vowel frequency, $F_p(1, 46) = 1.99$, *ns*, $F_i(1, 66) = 2.29$, *ns*, or consistency, $F_s < 1$, and the three-way interaction between these variables was also non-significant, $F_s < 1$. In the second analysis, accuracy was higher in the lexical phonology condition than in the no pre-exposure condition, $F_p(1, 46) = 13.06$, $p =$

.001, $F_i(1, 66) = 80.63, p < .001$. The effect of pre-exposure did not interact with vowel frequency, $F_s < 1$, or consistency, $F_p < 1, F_i(2, 66) = 1.52, ns$, and the three-way interaction between these variables was again non-significant, $F_s < 1$.

In both the above analyses the main effects of vowel frequency (defs vs. none: $F_p(1, 46) = 24.24, p < .001, F_i(1, 66) = 23.15, p < .001$, lex phon vs. none: $F_p(1, 46) = 33.29, p < .001, F_i(1, 66) = 31.48, p < .001$) and consistency (defs vs. none: $F_p(2, 46) = 8.02, p = .001, F_i(2, 66) = 8.28, p = .001$, lex phon vs. none: $F_p(2, 92) = 6.18, p < .01, F_i(2, 66) = 6.92, p < .01$) were significant and took the same form as reported in the preceding analysis. The interaction between frequency and consistency was not significant in either analysis (defs vs. none: $F_s < 1$, lex phon vs. none: $F_p(2, 92) = 1.11, ns, F_i < 1$).

To summarise, both lexical phonology and semantic pre-exposure facilitated reading accuracy at the beginning of training, regardless of vowel frequency or consistency. However, semantic pre-exposure was no more beneficial than pre-exposure to lexical phonology alone.

End of training

Lexical phonology vs. definitions. The mean number of blocks required to achieve the criterion level of $\geq 70\%$ accuracy was 1.97 ($SD = 1.06$). An analysis of variance examined the impact of pre-exposure on reading accuracy at the end of training. The results of this analysis are summarised in Figure 4 and Table 2. As in Experiment 1, this constituted different blocks for different participants. Accuracy was higher in the definitions condition than in the lexical phonology condition, $F_p(1, 31) = 6.76, p = .01, F_i(1, 66) = 7.73, p < .01$. The effect of pre-exposure interacted with frequency, $F_p(1, 31) = 5.38, p < .05, F_i(1, 66) = 5.37, p < .05$, and was significant in

items containing low but not high frequency vowels. Pre-exposure did not interact with consistency, $F_s < 1$. Although the three-way interaction between pre-exposure, frequency and consistency was not significant, $F_p < 1$, $F_i(2, 66) = 1.45$, *ns*, pairwise comparisons indicated that the advantage of definitions over lexical phonology pre-exposure in items containing low frequency vowels was only significant for inconsistent items.

The main effects of vowel frequency, $F_p(1, 31) = 43.72$, $p < .001$, $F_i(1, 66) = 41.96$, $p < .001$, and consistency, $F_p(2, 62) = 7.35$, $p = .001$, $F_i(2, 66) = 9.28$, $p < .001$, and the interaction between frequency and consistency, $F_p(2, 62) = 4.76$, $p = .01$, $F_i(2, 66) = 4.74$, $p = .01$, were significant and took the same form as reported for the end of training in Experiment 1.

INSERT TABLE 2 ABOUT HERE

Definitions/lexical phonology vs. no pre-exposure. Accuracy was marginally higher in the definitions condition than in the no pre-exposure condition, $F_p(1, 46) = 3.10$, $p = .09$, $F_i(1, 66) = 3.28$, $p = .08$. Across items (but not subjects) there were also marginal interactions between pre-exposure and vowel frequency, $F_p(1, 46) = 2.20$, $p = .15$, $F_i(1, 66) = 3.28$, $p = .08$, and consistency, $F_p(2, 92) = 1.61$, *ns*, $F_i(2, 66) = 2.65$, $p = .08$. The three-way interaction between pre-exposure, frequency and consistency was also significant by-items, $F_i(2, 66) = 3.29$, $p < .05$, although not by-subjects, $F_p(2, 92) = 1.83$, $p = .17$. Pairwise comparisons across both subjects and items demonstrated that within items containing low frequency inconsistent-unconditioned vowels, accuracy was higher in the definitions than the no pre-exposure condition.

In a second analysis which compared the lexical phonology and no pre-exposure conditions, the main effect of pre-exposure was not significant, $F_s < 1$. There was no interaction between pre-exposure and frequency, $F_s < 1$. The interaction between pre-exposure and consistency was significant by-items, $F_i(2, 66) = 3.09, p = .05$, but not by-subjects, $F_p(2, 92) = 1.41, ns$. Pairwise comparisons across items demonstrated that within consistent items accuracy was in fact lower following pre-exposure to lexical phonology than in the no pre-exposure condition and that pre-exposure had no effect on inconsistent items. There was no interaction between pre-exposure, frequency and consistency, $F_p < 1, F_i(2, 66) = 1.46, ns$.

In both of the above analyses, the main effects of frequency (defs vs. none: $F_p(1, 46) = 37.19, p < .001, F_i(1, 66) = 48.24, p < .001$, lex phon vs. none: $F_p(1, 46) = 58.35, p < .001, F_i(1, 66) = 58.56, p < .001$), and consistency (defs vs. none: $F_p(2, 92) = 16.31, p < .001, F_i(2, 66) = 23.39, p < .001$, lex phon vs. none: $F_p(2, 92) = 13.80, p < .001, F_i(2, 66) = 17.62, p < .001$), and the interaction between frequency and consistency (defs vs. none: $F_p(2, 92) = 3.61, p < .05, F_i(2, 66) = 5.66, p < .01$, lex phon vs. none: $F_p(2, 92) = 6.41, p < .01, F_i(2, 66) = 8.41, p = .001$) were significant and took the same form as reported for the end of training in Experiment 1.

To summarise the end of training performance, pre-exposure to semantics enhanced reading accuracy for items containing low frequency inconsistent vowels relative to both the lexical phonology and no pre-exposure conditions.

Old-new decision

Lexical phonology vs. definitions. We did not include frequency and consistency in old-new decision analyses because semantic effects do not interact with these variables in standard lexical decision. All data points were included for

analyses of discrimination accuracy. For latency analyses, only correct responses were included and trials with RTs more than two standard deviations away from each participant's mean were excluded (4.8%). One-sample t-tests confirmed that accuracy was above chance on both trained, $t_p(31) = 18.47, p < .001$, and untrained items, $t_p(31) = 8.84, p < .001$. Paired samples t-tests demonstrated that accuracy for trained, $M = .80 (.09)$, and untrained items, $M = .76 (.17)$, did not differ, $t_p(31) = 1.15, ns$. RTs were faster to trained, $M = 3481 (863)$, than untrained items, $M = 4389 (1435)$, $t_p(31) = 6.11, p < .001$.

The effect of pre-exposure on old-new discrimination accuracy was examined using d' . For each participant, we calculated the proportion of hits in each condition (lexical phonology $M = .74 (.13)$, definitions $M = .83 (.09)$). The mean proportion of false alarms, $M = .24 (.16)$, was calculated across conditions because untrained items did not receive lexical phonology or semantic pre-exposure. These values were transformed to z-scores and $z(\text{false alarms})$ was subtracted from $z(\text{lexical phonology hits})$ and $z(\text{definitions hits})$ to give a d' statistic for each participant in each condition. A paired samples by-subjects t-test indicated that d' was higher in the definitions condition, $M = 1.97 (1.29)$, than in the lexical phonology condition, $M = 1.56 (.90)$, $t_p(31) = 3.29, p < .01$, demonstrating that discrimination accuracy was enhanced by pre-exposure to definitions. Paired samples by-subject and by-items t-tests were then conducted on RTs to correct trained items. These demonstrated that pre-exposure did not affect trained item recognition latency (definitions, $M = 3447 (858)$, lexical phonology, $M = 3473 (948)$, $t_s < 1$).

Definitions/lexical phonology vs. no pre-exposure. Two sets of independent samples by-subjects t-tests comparing definitions/lexical phonology pre-exposure to

no pre-exposure were conducted on d' statistics and RTs to correct trained items. Items analyses were not possible due to the greater number of items included in Experiment 2 ($n = 36$) relative to Experiment 1 ($n = 12$). As reported in Experiment 1, in the no pre-exposure condition the mean proportions of hits and false alarms were .81 (.09) and .37 (.11) respectively, yielding a mean d' statistic of 1.26 (.44). The mean RT to correct trained items was 3775 (989). d' was higher in the definitions condition than in the no pre-exposure condition, $t_p(46) = 2.11, p < .05$. In contrast, d' was equivalent in the lexical phonology and no pre-exposure conditions, $t_p(46) = 1.26, ns$. RTs in the no pre-exposure condition did not differ from those in the definitions, $t_p(46) = 1.14, ns$, or lexical phonology condition, $t_p < 1$.

To summarise, old-new discrimination accuracy was facilitated by pre-exposure to definitions, relative to lexical phonology and no pre-exposure. Definitions pre-exposure improved discrimination by reducing the number of untrained items that were incorrectly accepted as targets, while preserving the number of trained items that were correctly recognised. In contrast, in the lexical phonology condition, the reduction in false positives also increased the number of trained items that were incorrectly rejected. Old-new decision latencies were unaffected by pre-exposure to lexical phonology or semantics.

Generalization

As the lexical phonology vs. semantics manipulation was within subject, we were unable to assess its effect on generalization. However, we could compare generalization following pre-exposure in Experiment 2 to the no pre-exposure baseline condition provided by Experiment 1, as summarised in Figure 5. There was a main effect of vowel frequency, $F_p(1, 46) = 27.68, p < .001, F_i(1, 18) = 77.06, p < .001$.

Accuracy was higher for items containing consistent vowels than for those containing inconsistent-conditioned vowels, which in turn outperformed items containing inconsistent-unconditioned vowels, $F_p(2, 92) = 18.13, p < .001, F_i(2, 18) = 53.02, p < .001$. Unlike in Experiment 1, the interaction between frequency and consistency reached significance, $F_p(2, 92) = 10.99, p < .001, F_i(2, 18) = 18.11, p < .001$. Pairwise comparisons indicated that the frequency effect was only significant in items containing inconsistent vowels and that the consistency effect was only significant in items containing low frequency vowels.

Generalization was marginally lower in Experiment 2 than in Experiment 1, $F_p(1, 46) = 3.65, p = .06, F_i(1, 18) = 3.50, p = .08$. This effect did not interact with vowel frequency, $F_s < 1$, or consistency, $F_p(2, 92) = 1.13, ns, F_i(2, 18) = 1.14, ns$. However, the three-way interaction between these variables was significant by-subjects, $F_p(2, 92) = 3.18, p < .05$, although not by-items, $F_i(2, 18) = 1.82, ns$. Pairwise comparisons across subjects indicated that the lower accuracy in Experiment 2 was only significant in items containing low frequency, inconsistent vowels. This can be clearly seen in Figure 5.

INSERT FIGURE 5 ABOUT HERE

Discussion

Experiment 2 used the artificial orthography paradigm to examine the influence of familiarity with word meanings vs. familiarity with word sounds on orthographic learning. Participants were pre-exposed to either the phonological forms of items or to their phonological form plus a definition and we examined subsequent orthographic learning, discrimination and generalization.

We expected semantic pre-exposure to facilitate reading accuracy at the end of training and old-new decision, relative to both lexical phonology and no pre-exposure. In reading aloud we expected facilitation to be restricted to items containing low frequency inconsistent vowels. These predictions were confirmed; additionally, we found that pre-exposure to lexical phonology provided no observable benefits, relative to the no pre-exposure condition of Experiment 1. These findings provide convincing evidence that semantic knowledge supports word reading processes, in line with the predictions of the triangle model and the existing adult data (Balota et al., 2004; Cortese & Khanna, 2007; McKay et al., 2007; Patterson et al., 2006; Pexman et al., 2008; Strain et al., 1995; Woollams, 2005; Woollams et al., 2007).

A second prediction was that early in training, phonological familiarity would be of primary importance with semantic pre-exposure providing no additional benefit. At this stage it was also possible that the benefits of pre-exposure would be pervasive rather than restricted to particular item types. Again, the data supported our predictions. These results concord with developmental evidence (McKague et al., 2001; Nation & Cocksey, 2009) and suggest that the early stages of orthographic learning may be more dependent on phonological than semantic familiarity.

Turning to generalization, novel word reading was poorer in Experiment 2 than in Experiment 1, particularly when items contained low frequency and inconsistent vowels. We suggest two possible explanations for this finding, both of which are elaborated on in the General Discussion. First, pre-exposure increased knowledge of how the training items should sound. This may have enabled participants to guess at how they should be pronounced during orthographic

learning. Second, pre-exposure may have increased the extent to which training items were treated as discrete entities. Both these possibilities could reduce the information participants extracted about conditional spelling-sound mappings.

General Discussion

The artificial orthography paradigm allowed us to investigate a number of questions about orthographic learning. First, we asked whether learners are able to extract sub-word spelling-sound patterns from exposure to the orthographic and phonological forms of whole words, without explicit instruction. The results of both experiments show that they can. Second, and related to this, learners were sensitive to two lexical properties inherent in the training set; items containing high frequency and high consistency vowels were learned more easily. Third, we asked whether learners can generalize their knowledge to novel forms. They could, and generalization in both experiments was influenced by vowel frequency and consistency statistics extracted from exposure. Finally, Experiment 2 found that familiarising participants to the sounds (lexical phonology) or sounds plus meanings of words (semantics) influenced learning and generalisation, with semantic information emerging as a stronger influence on learning as training proceeded.

Taken together, these findings highlight the utility of the artificial orthography paradigm. Participants were sensitive to statistical patterns (frequency and consistency) embedded within the novel orthography. Furthermore, their sensitivity resembled effects observed in natural languages: orthography-phonology inconsistency was less detrimental when a pattern was experienced often. In our

view, this suggests that the paradigm has good validity for exploring theoretical questions concerning orthographic learning.

Before discussing some of the theoretical implications of our findings, one potential concern must be discussed. As our participants were highly literate adults, perhaps they were simply mapping the artificial orthography onto their knowledge of English letter-sound mappings? This seems unlikely, especially for vowels: the vowel sounds in our orthography are represented in multiple ways in written English. For example, the phoneme /i/ can be spelled, E, EE, EA, IE, EI, EY, and /u/ can be spelled, U, O, OO, UE, OU, UI, EW. Therefore, vowel character-phoneme relationships do not directly map to English. Another concern might be that the context-sensitive rules that governed inconsistent vowel pronunciations in the artificial orthography were deterministic. Arguably, these might be easier to abstract than the probabilistic conditional relationships that exist in English (Treiman et al., 1995). However, the less than perfect generalization speaks against this. In addition, post-experiment debriefing revealed that only three participants (from a total of 64 across experiments) were explicitly aware of any of the conditional consonant-vowel pronunciation rules. In sum, learning the novel orthography was achievable but not trivial, and most participants were not able to identify the rules that they had nevertheless extracted.

Extracting spelling-sound regularities

When children first come to the task of learning to read they do not know which words follow regular spelling-sound patterns and which do not. We might assume this knowledge emerges over time through experience with reading words, but previous work contradicts this suggesting that adults are unable to extract and

generalize sub-word regularities when learning an artificial orthography (Bitan & Karni, 2003;2004). However, as discussed earlier, methodological issues limit the conclusions that can be drawn from these studies. Our work therefore makes an important theoretical contribution by providing the first demonstration that learners can abstract sub-word regularities (and use them in generalization) through exposure to only whole-word orthography-phonology mappings and without explicit instruction.

Moreover, learners' sensitivity tracked the statistics of the orthography impressively well: consistent items were learned more easily than inconsistent items and, within these, conditioned vowels were learned more easily than unconditioned vowels. This provides direct support for Jared and colleagues' (Jared, 1997;2002; Jared et al., 1990) work which suggested that knowledge of spelling-sound regularity is graded. It also furthers Kessler and Treiman's (2001) work on the influence of consonant context on vowel pronunciation predictability. We found that vowels with more predictable pronunciations were easier to learn, that their pronunciations were extracted for use in generalization, and that these effects were modulated by how frequently vowels had been experienced.

The effect of pre-exposure on orthographic learning

Experiment 2 found that by the end of training, reading accuracy was enhanced by pre-exposure to item definitions but not by pre-exposure to item sounds. Furthermore, this semantic benefit was specific to items containing low frequency inconsistent vowels. As our data are not compromised by problems in measuring and controlling for the many variables that influence word reading, or by

difficulties in pulling apart effects of semantic and phonological familiarity, our results provide support for the role of semantics in reading aloud.

An interesting developmental pattern emerged when we considered reading success at an earlier stage of training. In Block 1, pre-exposure to item sounds boosted reading accuracy relative to no pre-exposure and semantic pre-exposure provided no additional benefit. Furthermore, the benefit conferred by phonological familiarity early in training was not restricted to items containing inconsistent low frequency vowels; instead, facilitation was seen across item types. These results are in line with data from children showing no semantic advantage in beginning readers (McKague et al., 2001; Nation & Cocksey, 2009), beyond the benefits provided by familiarity with lexical phonology. Previous word learning experiments with adults that have used English orthography have not been able to consider such early stages of learning because participants have highly proficient knowledge of their own orthography (c.f. Bowers et al., 2005; McKay et al., 2008). Overall, our findings highlight the importance of considering learning if we are to gain a comprehensive picture of the role of semantics in word reading processes.

Our observation that semantic knowledge begins to contribute to reading aloud later in learning, perhaps once orthography-phonology mappings have developed to some degree, is consistent with the division of labour hypothesis that emerged from Plaut et al.'s (1996) triangle model. They argued that semantic effects arise later "in part because of the phonological nature of typical reading instruction and in part because in English, the orthography-to-phonology mapping is far more structured than the orthography-to-semantics mapping" (p. 95, Plaut et al., 1996).

Is our finding that item-specific knowledge of word sounds facilitated performance in the early stages of learning also compatible with the triangle model framework? Although *lexical* phonology is not represented in the model, a later version did incorporate a phonological attractor network which increased the model's "knowledge of the segmental structure and constraints on sequences of phonemes" (Harm & Seidenberg, 1999, p. 43). This provides a way for phonological familiarity (although not necessarily at a whole-word level) to influence learning of orthography-phonology mappings. In addition, a reconsideration of Plaut et al.'s (1996) implementation of the triangle model suggests an alternative way in which whole-word phonological input could be provided. In their implementation, 'semantic support' operated by providing "additional input to the phonological units, pushing them towards their correct activations" (p. 95). This input could quite easily be re-cast as arising from the possession of item-specific (or lexical) phonological representations. Future modelling work should examine whether a mechanism that encourages whole-word phonological representations facilitates the early stages of learning to read.

One point that should be revisited here is that in the triangle model, semantic knowledge is most important for reading inconsistent *words* that occur infrequently in the language. In contrast, in our paradigm we obtained an influence of semantics on items containing inconsistent *vowels* that were low frequency. Whilst both vowel and word frequency influence the strength of orthography-phonology mappings, word frequency also allows a system to overcome sublexical inconsistencies by placing greater weight on larger (e.g. whole-word) units. Furthermore, word frequency impacts on the bonds between orthography/phonology and semantics,

whereas this is not true of vowel frequency.¹ Future work should therefore examine the impact of semantics on artificial orthography learning when a word token frequency manipulation is incorporated.

Pre-exposure effects on discrimination

So far, this discussion has focussed on how well learners were able to read the training items. We also assessed their ability to discriminate trained words from novel words written in the same orthography, using an old-new decision task. Although discrimination was above chance, participants made a number of errors and showed very long RTs. Interestingly however, pre-exposure to semantic information enhanced discrimination in Experiment 2, relative to both no pre-exposure (Experiment 1) and pre-exposure to lexical phonology. This suggests that meaning increased participants' certainty of which items they did and did not know. By what mechanism did prior knowledge of word meanings aid recognition? Patterson and Hodges (1992) noted that the phonological representations of familiar words differ from those of unfamiliar words in two important ways. First, the sounds in known words co-occur more often than the sounds in novel words. Thus, we might expect pre-exposure to lexical phonology (which increased familiarity with the sounds within training items) to improve discrimination. Second, the combination of sounds in a known word corresponds to a specific meaning, unlike the sounds in a novel word. We would therefore expect facilitation to be even greater following semantic pre-exposure. In fact, discrimination was only enhanced in the semantic condition.

We suggest two possible reasons why this might be the case. The first is that greater experience with (or attention to) item sounds might be necessary for lexical

phonology pre-exposure to produce significant benefit. Participants simply had to listen to and repeat items and there was no pressure to commit them to memory. A second possibility is that “meaning is the source of the glue that holds the phonological elements of a word together” (Patterson & Hodges, 1992, p. 1036). On this view, pre-exposure to word meanings was beneficial not only because it provided a semantic association, but also because it boosted learning about phonological co-occurrence, highlighting item-specificity.

It is worth noting at this point that although pre-exposure to semantics increased discrimination accuracy, it did not influence decision latencies. Overall, responses were slow in both experiments, in the order of seconds rather than milliseconds. This indicates that item-specific orthographic knowledge was not immediately available (cf. typical lexical decision). An important question concerns how the representations of these newly learned words differ from those of words in natural languages. One possibility is that they are episodic rather than lexical. This seems likely given the rather short exposure phase followed immediately by the test phase. Future work could examine the lexical basis of newly learned words using markers of lexical processing such as masked priming (Bowers, 2003; Forster, Mohan, & Hector, 2003) or the prime-lexicality effect (Davis & Lupker, 2006). Lexicalisation during orthographic learning could also be examined by assessing whether newly learned forms engage in competition with existing words, drawing on recent work in spoken word recognition (Dumay & Gaskell, 2007; Gaskell & Dumay, 2003).

Pre-exposure effects on generalization

Our final point to note is that although pre-exposure to semantics improved discrimination, it brought with it a cost: generalization was less good following pre-exposure than no pre-exposure, and this decline was particularly pronounced for items containing low frequency and inconsistent vowels. Two non-exclusive reasons for this seem plausible. First, pre-exposure increased knowledge of how trained items sounded, allowing participants to guess how they should be pronounced during orthographic learning. This would reduce the necessity to abstract information about the influence of consonants on vowel pronunciation. A related possibility is that prior knowledge of item sounds or meanings increased the extent to which items were treated as discrete entities. If items were regarded as more different from each other, this might also reduce the abstraction of conditional spelling-sound rules for use in generalization.

Consistent with the view that any manipulation that reduces attention to character-sound mappings might be detrimental for generalization, Harm, McCandliss and Seidenberg (2003) found that, in a connectionist model, training which emphasised sub-word orthography-phonology regularities improved generalization relative to training which focused on phonology alone. Harm and Seidenberg (1999) provide some insight into why this might be the case. They found that a connectionist model that formed less componential and more holistic orthography-phonology mappings performed poorly in generalization. Holistic representations of similar words (e.g., MEAT, TREAT and EAT) were less overlapping and this reduced the model's ability to read a novel word such as GEAT because "the pattern of hidden unit activity generated by GEAT was not close enough to the representation of similar words." (p. 173). Although this observation demonstrates

why abstraction of conditional character-sound mappings is important, it does not provide strong evidence for either of our suggestions as to why pre-exposure inhibited this process in our experiments.

More direct evidence is provided by Landi, Perfetti, Bolger, Dunlap and Foorman (2006). They found that children were better at learning to read new words presented in connected text than new words presented in isolation. However, when re-tested one week later, reading accuracy was higher for those words learned in isolation. This suggests that although context increased the chance of reading words correctly during training, it reduced attention to orthography-phonology mappings which in turn compromised longer term retention. Landi et al. also noted that learning in context had a less negative effect on retention for those children who were more advanced readers. This suggests that pre-exposure would be less detrimental for generalization if orthographic training was more extensive. These findings support our first proposal that pre-exposure provided an additional source of information to support reading acquisition, simultaneously reducing the extraction of conditional character-sound mappings. However, they do not rule out our second proposal that pre-exposure made learning more item-specific.

Conclusions

The experiments reported here have shown that an artificial orthography paradigm can be used to investigate learning and generalization in reading aloud. We found that adults can extract sub-word regularities from exposure to whole word orthographic forms and their pronunciations, and that they can use these productively. Vowels frequency and consistency influenced learning in predictable ways showing that learners are sensitive to lexical properties that are implicit in the

language environment. Experiment 2 demonstrated how the paradigm can be used to address a theoretical question - in our case concerning the role of semantics in orthographic learning. Our experiments are in line with previous research suggesting that semantic knowledge supports word reading processes. However, we also demonstrated that familiarity with an item's phonological form plays a role earlier in development, highlighting the need to consider learning and developmental change.

References

- Andrews, S. (1982). Phonological recoding: Is the regularity effect consistent. *Memory & Cognition, 10*(6), 565-575.
- Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General, 133*(2), 283-316.
- Bitan, T., & Karni, A. (2003). Alphabetical knowledge from whole words training: Effects of explicit instruction and implicit experience on learning script segmentation. *Cognitive Brain Research, 16*(3), 323-337.
- Bitan, T., & Karni, A. (2004). Procedural and declarative knowledge of word recognition and letter decoding in reading an artificial script. *Cognitive Brain Research, 19*(3), 229-243.
- Blazely, A. M., Coltheart, M., & Casey, B. J. (2005). Semantic impairment with and without surface dyslexia: Implications for models of reading. *Cognitive Neuropsychology, 22*, 695-717.

- Bonin, P., Barry, C., Meot, A., & Chalard, M. (2004). The influence of age of acquisition in word reading and other tasks: A never ending story? *Journal of Memory and Language, 50*(4), 456-476.
- Bowers, J. S. (2003). An abstractionist account of masked and long-term priming. In S. Kinoshita & S. J. Lupker (Eds.), *Masked Priming: The State of the Art* (pp. 39-55). New York and Hove: Psychology Press.
- Bowers, J. S., Davis, C. J., & Hanley, D. A. (2005). Interfering neighbours: The impact of novel word learning on the identification of visually similar words. *Cognition, 97*(3), B45-B54.
- Brown, G. D. A., & Watson, F. L. (1987). 1st in, 1st out - Word Learning Age and Spoken Word-Frequency as Predictors of Word Familiarity and Word Naming Latency. *Memory & Cognition, 15*(3), 208-216.
- Byrne, B. (1984). On teaching articulatory phonetics via an orthography. *Memory & Cognition, 12*(2), 181-189.
- Byrne, B., & Carroll, M. (1989). Learning artificial orthographies - Further evidence of a nonanalytic acquisition procedure. *Memory & Cognition, 17*(3), 311-317.
- Callan, A. M., Callan, D. E., & Masaki, S. (2005). When meaningless symbols become letters: Neural activity change in learning new phonograms. *NeuroImage, 28*(3), 553-562.
- Coltheart, M. (2005). Modelling reading: The dual route approach. In M. Snowling & C. Hulme (Eds.), *The Science of Reading: A Handbook*. Oxford: Blackwell Publishing.

- Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud - Dual-route and parallel-distributed-processing approaches. *Psychological Review*, *100*(4), 589-608.
- 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*(1), 204-256.
- Cortese, M. J., & Khanna, M. M. (2007). Age of acquisition predicts naming and lexical-decision performance above and beyond 22 other predictor variables: An analysis of 2,342 words. *Quarterly Journal of Experimental Psychology*, *60*(8), 1072-1082.
- Cortese, M. J., & Simpson, G. B. (2000). Regularity effects in word naming: What are they? *Memory & Cognition*, *28*(8), 1269-1276.
- Davis, C. J., & Lupker, S. J. (2006). Masked inhibitory priming in English: Evidence for lexical inhibition. *Journal of Experimental Psychology-Human Perception and Performance*, *32*(3), 668-687.
- Dumay, N., & Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science*, *18*(1), 35-39.
- Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture - A critical analysis. *Cognition*, *28*(1-2), 3-71.
- Forster, K., Mohan, K., & Hector, J. (2003). The mechanics of masked priming. In S. Kinoshita & S. J. Lupker (Eds.), *Masked Priming: The State of the Art* (pp. 3-37). New York and Hove: Psychology Press.
- Forster, K. I. (2000). The potential for experimenter bias effects in word recognition experiments. *Memory & Cognition*, *28*(7), 1109-1115.

Frost, S. J., Mencl, W. E., Sandak, R., Moore, D. L., Rueckl, J. G., Katz, L., et al. (2005).

A functional magnetic resonance imaging study of the tradeoff between semantics and phonology in reading aloud. *Neuroreport*, *16*(6), 621-624.

Gaskell, M. G., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition*, *89*(2), 105-132.

Gilhooly, K., & Watson, F. (1981). Word age-of-acquisition effects: a review. *Current Psychology*, *1*(3), 269-286.

Glushko, R. J. (1979). Organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, *5*(4), 674-691.

Harm, M. W., McCandliss, B. D., & Seidenberg, M. S. (2003). Modeling the successes and failures of interventions for disabled readers. *Scientific Studies of Reading*, *7*(2), 155-182.

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review*, *106*(3), 491-528.

Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review*, *111*(3), 662-720.

Hashimoto, R., & Sakai, K. L. (2004). Learning letters in adulthood: Direct visualization of cortical plasticity for forming a new link between orthography and phonology. *Neuron*, *42*(2), 311-322.

- Humphreys, G. W., & Evett, L. J. (1985). Are there independent lexical and nonlexical routes in word-processing? An evaluation of the dual-route theory of reading. *Behavioral and Brain Sciences*, *8*(4), 689-705.
- Jacobs, A. M., & Grainger, J. (1994). Models of visual word recognition - Sampling the state-of-the-art. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(6), 1311-1334.
- Jared, D. (1997). Spelling-sound consistency affects the naming of high-frequency words. *Journal of Memory and Language*, *36*(4), 505-529.
- Jared, D. (2002). Spelling-sound consistency and regularity effects in word naming. *Journal of Memory and Language*, *46*(4), 723-750.
- Jared, D., McRae, K., & Seidenberg, M. S. (1990). The basis of consistency effects in word naming. *Journal of Memory and Language*, *29*(6), 687.
- Kessler, B., & Treiman, R. (2001). Relationships between sounds and letters in English monosyllables. *Journal of Memory and Language*, *44*(4), 592-617.
- Kessler, B., Treiman, R., & Mullenix, J. (2003). A new look at the factors that affect oral word reading. Unpublished manuscript. .
- Knafle, J. D., & Legenza, A. (1978). Generalizability of inquiry involving artificial orthography. *American Educational Research Journal*, *15*(2), 331-347.
- Lacruz, I., & Folk, J. R. (2004). Feedforward and feedback consistency effects for high- and low-frequency words in lexical decision and naming. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *57*(7), 1261-1284.
- Landi, N., Perfetti, C. A., Bolger, D. J., Dunlap, S., & Foorman, B. R. (2006). The role of discourse context in developing word form representations: A paradoxical

- relation between reading and learning. *Journal of Experimental Child Psychology*, 94(2), 114-133.
- Magnuson, J. S. (2008). Nondeterminism, pleiotropy, and single-word reading: theoretical and practical concerns. In E. L. Grigorenko & A. J. Naples (Eds.), *Single-Word Reading: Behavioral and Biological Perspectives*. New York, London: Lawrence Erlbaum Associates.
- Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (2003). The time course of spoken word learning and recognition: Studies with artificial lexicons. *Journal of Experimental Psychology: General*, 132(2), 202-227.
- McCandliss, B. D., Posner, M. I., & Givon, T. (1997). Brain plasticity in learning visual words. *Cognitive Psychology*, 33(1), 88-110.
- McKague, M., Pratt, C., & Johnston, M. B. (2001). The effect of oral vocabulary on reading visually novel words: a comparison of the dual-route-cascaded and triangle frameworks. *Cognition*, 80(3), 231-262.
- McKay, A., Castles, A., Davis, C., & Savage, G. (2007). The impact of progressive semantic loss on reading aloud. *Cognitive Neuropsychology*, 24(2), 162-186.
- McKay, A., Davis, C., Savage, G., & Castles, A. (2008). Semantic involvement in reading aloud: Evidence from a nonword training study. *Journal of Experimental Psychology: Learning Memory and Cognition*, 34(6), 1495-1517.
- Monaghan, J., & Ellis, A. W. (2002). What exactly interacts with spelling-sound consistency in word naming? *Journal of Experimental Psychology: Learning Memory and Cognition*, 28(1), 183-206.

- Morrison, C. M., & Ellis, A. W. (1995). Roles of word-frequency and age of acquisition in word naming and lexical decision. *Journal of Experimental Psychology: Learning Memory and Cognition*, *21*(1), 116-133.
- Morrison, C. M., & Ellis, A. W. (2000). Real age of acquisition effects in word naming and lexical decision. *British Journal of Psychology*, *91*, 167-180.
- Morrison, C. M., Hirsh, K. W., Chappell, T., & Ellis, A. W. (2002). Age and age of acquisition: An evaluation of the cumulative frequency hypothesis. *European Journal of Cognitive Psychology*, *14*(4), 435 - 459.
- Nation, K., & Cocksey, J. (2009). The relationship between knowing a word and reading it aloud in children's word reading development. *Journal of Experimental Child Psychology*, *103*(3), 296-308.
- Patterson, K., & Hodges, J. R. (1992). Deterioration of word meaning: Implications for reading. *Neuropsychologia*, *30*(12), 1025-1040.
- Patterson, K., Lambon-Ralph, M. A., Jefferies, E., Woollams, A., Jones, R., Hodges, J. R., et al. (2006). "Presemantic" cognition in semantic dementia: Six deficits in search of an explanation. *Journal of Cognitive Neuroscience*, *18*(2), 169-183.
- Pexman, P. M., Hargreaves, I. S., Siakaluk, P. D., Bodner, G. E., & Pope, J. (2008). There are many ways to be rich: Effects of three measures of semantic richness on visual word recognition. *Psychonomic Bulletin & Review*, *15*(1), 161-167.
- Pinker, S. (1991). Rules of language. *Science*, *253*(5019), 530-535.
- Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, *103*(1), 56-115.

- Schwartz, M. F., Marin, O. S. M., & Saffran, E. M. (1979). Dissociations of language function in dementia: A case study. *Brain and Language*, *7*(3), 277-306.
- Seidenberg, M. S. (1985). The time course of phonological code activation in 2 writing systems. *Cognition*, *19*(1), 1-30.
- Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, *96*(4), 523-568.
- Seidenberg, M. S., Waters, G. S., Barnes, M. A., & Tanenhaus, M. K. (1984). When does irregular spelling or pronunciation influence word recognition. *Journal of Verbal Learning and Verbal Behavior*, *23*(3), 383-404.
- Stone, G. O., Vanhoy, M., & VanOrden, G. C. (1997). Perception is a two-way street: Feedforward and feedback phonology in visual word recognition. *Journal of Memory and Language*, *36*(3), 337-359.
- Strain, E., Patterson, K., & Seidenberg, M. S. (1995). Semantic effects in single-word naming. *Journal of Experimental Psychology: Learning Memory and Cognition*, *21*(5), 1140-1154.
- Taraban, R., & McClelland, J. L. (1987). Conspiracy effects in word pronunciation. *Journal of Memory and Language*, *26*(6), 608-631.
- Treiman, R., Kessler, B., & Bick, S. (2003). Influence of consonantal context on the pronunciation of vowels: A comparison of human readers and computational models. *Cognition*, *88*(1), 49-78.
- Treiman, R., Mullennix, J., Bijeljacabic, 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*(2), 107-136.

Waters, G. S., & Seidenberg, M. S. (1985). Spelling-sound effects in reading - Time-course and decision criteria. *Memory & Cognition*, *13*(6), 557-572.

Wonnacott, E., Newport, E. L., & Tanenhaus, M. K. (2008). Acquiring and processing verb argument structure: Distributional learning in a miniature language. *Cognitive Psychology*, *56*(3), 165-209.

Woollams, A. M. (2005). Imageability and ambiguity effects in speeded naming: Convergence and divergence. *Journal of Experimental Psychology: Learning Memory and Cognition*, *31*(5), 878-890.

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

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*(4), 1131-1161.

Appendix A





Novel Characters and Phoneme Mappings

Character	Phoneme Mapping 1	Phoneme Mapping 2
Λ	/u/ or /ʌ/	/aɪ/ or /ɪ/
‡	/b/	/f/
χ	/p/	/m/
ℓ	/i/ or /ɛ/	/əʊ/ or /ɒ/
⊗	/əʊ/ or /ɒ/	/i/ or /ɛ/
ƒ	/d/	/b/
ƒ	/s/	/n/
ε	/m/	/z/
⊙	/f/	/g/
≡	/aɪ/ or /ɪ/	/u/ or /ʌ/
◻	/v/	/s/
◻	/n/	/v/
⋈	/z/	/p/
√	/g/	/d/
ƒ	/t/	/k/
⊙	/k/	/t/

Appendix B

Frequency x Consistency Manipulation for training set 1, character-phoneme

mapping 1, as described in Experiment 1

Character	Pronunciation	Consistency	Onset	Frequency
	/u/	Consistent	Any	8 items
	/aɪ/	Consistent	Any	4 items
	/əʊ/	Inconsistent- conditioned	/z/ character	4 items
	/ɒ/	Inconsistent- unconditioned	any except /z/ character	8 items
	/i/	Inconsistent- conditioned	/v/ character	8 items
	/ɛ/	Inconsistent- unconditioned	any except /v/ character	4 items

Appendix C

Examples of Training and Test Sets

Table C1

Item Pronunciations in Training Set 1

High Frequency			Low Frequency		
Cons	Incons-cond	Incons-uncond	Cons	Incons-cond	Incons-uncond
buv, dus,	vid, vif, vig,	bɔz, gɔm, mɔn,	kaɪd, saɪv,	zəʊb, zəʊm,	tɛz, mɛp, zɛk,
fun, tup,	vik, vim, vip,	pɔs, dɔz, vɔb,	zaɪk, gaɪn	zəʊt, zəʊv	sɛv
kuf, nut,	vis, viv	fɔt, nɔf			
pug, vud					

Table C2

Item Pronunciations in Test Set 1

High Frequency			Low Frequency		
Cons	Incons-cond	Incons-uncond	Cons	Incons-cond	Incons-uncond
fub, mup	viz, vin	vɔn, pɔf	baɪm, zaɪg	zəʊd, zəʊs	gɛm, nɛv

Note. Frequency and consistency refer to the characteristics of vowels contained in test items during training

Footnotes

1. We would like to thank an anonymous reviewer for highlighting these differences between vowel and word frequency manipulations.

Table 1

Proportion of items correct (*SD*) at the Beginning of Training in Experiments 1 and 2, broken down according to Pre-Exposure and Item type.

		Low Frequency			High Frequency		
		Cons	Incons- Cond	Incons- Uncond	Cons	Incons- Cond	Incons- Uncond
Exp 1	None	.19 (.28)	.25(.30)	.06 (.11)	.38 (.34)	.51 (.22)	.27 (.24)
Exp 2	Lexical Phonology	.52 (.45)	.45 (.37)	.36 (.39)	.59 (.37)	.68 (.31)	.58 (.32)
	Semantics	.55 (.39)	.52 (.35)	.45 (.45)	.63 (.37)	.72 (.34)	.52 (.31)

Table 2

Proportion of items correct (*SD*) at the End of Training in Experiments 1 and 2,
broken down according to Pre-Exposure and Item type.

		Low Frequency			High Frequency		
		Cons	Incons- Cond	Incons- Uncond	Cons	Incons- Cond	Incons- Uncond
Exp 1	None	.86 (.27)	.64 (.34)	.38 (.26)	.93 (.09)	.86 (.11)	.80 (.27)
Exp 2	Lexical Phonology	.78 (.33)	.55 (.39)	.48 (.37)	.85 (.17)	.88 (.18)	.83 (.20)
	Semantics	.84 (.27)	.67 (.35)	.64 (.36)	.91 (.16)	.90 (.17)	.79 (.22)

Figure Captions

Figure 1. Proportion of items read correctly in each block of training in Experiment 1. Vertical lines depict standard deviations of subject means (*SD*). Numbers above the curve indicate the cumulative number of participants who had achieved 70%.

Figure 2. Proportion of items read correctly at the end of training in Experiment 1 as a function of vowel frequency and consistency. Vertical lines depict standard errors of subject means (*SE*).

Figure 3. Proportion of generalization items read correctly ($\pm 1SE$) in Experiment 1 as a function of the frequency and consistency of vowels during training.

Figure 4. Accuracy at the beginning and end of training ($\pm 1SE$) following semantics and lexical phonology pre-exposure in Experiment 2 and no pre-exposure in Experiment 1, as a function of frequency.

Figure 5. Generalization accuracy ($\pm 1SE$) in Experiments 1 and 2 as a function of frequency and consistency.

Figure 1

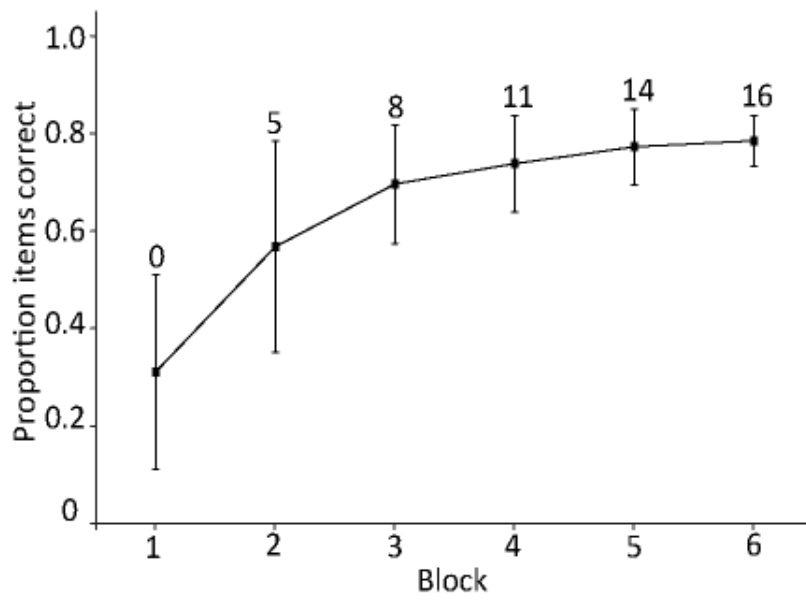


Figure 2

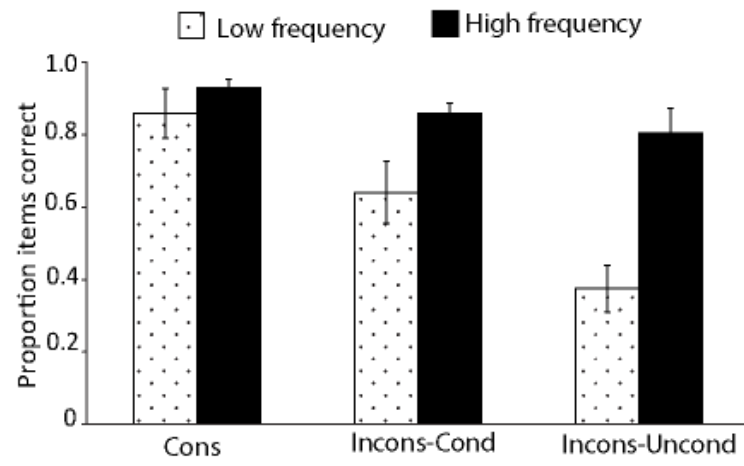


Figure 3

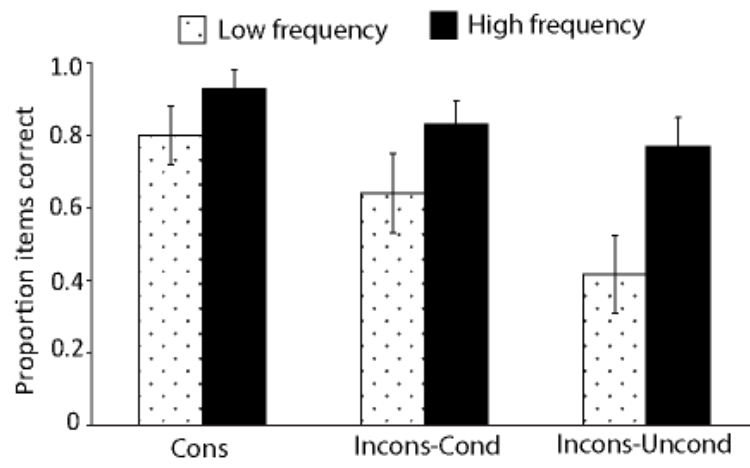


Figure 4

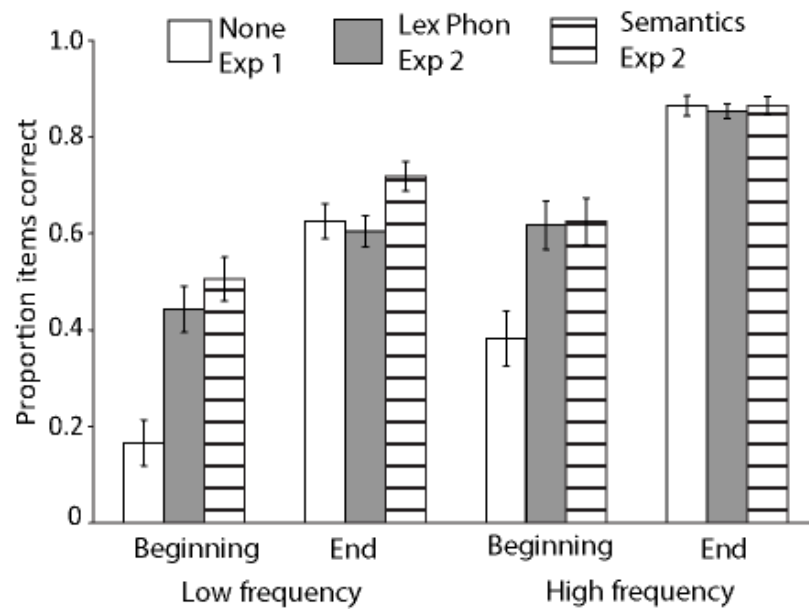


Figure 5

