

**Statistical learning and spelling: Evidence from an incidental learning experiment with children**

Anna Samara<sup>1</sup>, Daniela Singh<sup>1</sup>, & Elizabeth Wonnacott<sup>1</sup>

Division of Psychology and Language Sciences, University College London

Correspondence concerning this article should be addressed to Anna Samara, Division of Psychology and Language Sciences, University College London, Chandler House, 2 Wakefield Street, London, WC1N 1PF, United Kingdom

Email: [anna.samara@ucl.ac.uk](mailto:anna.samara@ucl.ac.uk)

**Declarations of interest: none**

**Word count = 2998**

### Abstract

Statistical learning processes—akin to those seen in spoken language acquisition (Saffran, Aslin, & Newport, 1996)—may be important for the development of literacy, particularly spelling development. One previous study provides direct evidence for this process: Samara and Caravolas (2014) demonstrated that 7-year-olds generalize over permissible letter contexts (graphotactics) in novel word-like stimuli under incidental learning conditions. However, unlike in actual orthography, conditioning contexts in Samara and Caravolas' (2014) stimuli comprised perfectly correlated, redundant cues in both *word-initial* and *word-final* positions. The current study explores whether 7-year-olds can extract such constraints in the absence of redundant cues. Since theories of literacy development predict greater sensitivity to restrictions within word-final units, we also contrast learning in word-initial and word-final units. We demonstrate that—for 7-year-old learners in two linguistic contexts (English and Turkish)—there is substantial evidence for the learning of both types of restriction.

*Keywords:* Statistical learning; spelling; graphotactic restrictions; incidental learning; word-final units; Bayes Factors

Abstract word count = 144

**Word count = 2998**

## 1. Introduction

Many empirical studies with infant and adult learners have established that statistical learning processes operate at multiple levels of spoken language (e.g., phonology, morphology, syntax) acquisition. Written language is another statistically patterned domain of knowledge, yet little work has directly assessed whether the same learning mechanisms are at play during spelling development, and how these are constrained. We report on a learning experiment with English- and Turkish-speaking children that addresses these questions.

Computational analyses of the English orthography has revealed a range of probabilistic rules that constrain the use of different graphemes in particular positions and contexts (Kessler & Treiman, 2001). Importantly, children are sensitive to such constraints. For example, Treiman and Kessler (2006) showed that 11-year-olds, asked to spell nonwords, were more likely to spell /ɛ/ followed by /d/ as “ea” (e.g., /glɛd/ → glead) as opposed to /ɛ/ followed by other codas (/glɛp/ → glep); eight-year-olds were more likely to spell /ɑ/ as “o” when preceded by the onset /w/ (e.g., /kwɑp/ → quap) as opposed to other onsets (e.g., /l/) (e.g., /blɑp/ → blɒp). These results suggest that children show sensitivity to contingencies between vowel spellings and the adjacent following/preceding consonants, and similar findings are seen in nonword judgments, in children’s own spelling errors, and for different type of constraints (e.g., purely graphotactic rules where conditioning has no phonological counterpart) (Cassar & Treiman, 1997; Hayes, Treiman, & Kessler, 2006; Pacton, Perruchet, Fayol, & Cleeremans, 2001; Pacton, Sobaco, Fayol, & Treiman, 2013; Treiman, 1993).

The key premise of these studies is that pattern knowledge develops from text exposure via statistical learning. Samara and Caravolas (2014) directly tested this among 7.5-year-olds building on work by Onishi, Chambers & Fisher (2002) in the phonotactic domain. They assessed learning of graphotactic “rules” that resembled those encountered in written English (e.g., “g and z cannot co-occur”) but were novel in nature (e.g., “o and p cannot co-occur”). The incidental learners saw Consonant-Vowel-Consonant letter strings while performing a cover (color detection) task. Unbeknown to them, there were restrictions between consonants and the neighbouring vowel *both* word initially (e.g., medial *o* was always preceded by two out of four consonants such that, for example, strings could not begin with *po*), *and* word finally (e.g., medial *o* was also followed by only two out of four consonants such that, for example, strings could not end with *ol*). At test, children discriminated “permissible” from “impermissible” novel stimuli suggesting learning and generalization over the novel restrictions without explicit instruction.

Samara and Caravolas (2014) provide strong first evidence that 7-year-olds rapidly extract graphotactic restrictions using similar processes to those at work in spoken language acquisition. This challenges popular models of literacy development, which propose that sensitivity to spelling emerges “late” (Frith, 1985; Marsh, Friedman, Welch, & Desberg, 1980). However, stimuli were designed to maximize cues available to the learner: vowels were cued by both preceding and following context, whereas earlier work (e.g., Treiman & Kessler, 2006) has investigated children’s sensitivity to *each* cue in isolation. Disentangling learning from preceding versus following context is particularly important given a long-standing debate regarding the relative importance of word-initial and word-final units in literacy development. One view (Fudge, 1969, 1987; Selkirk, 1982, Treiman, 1986; Treiman, Mullennix, Bijeljac-Babic, & Richmond-Welty, 1995), is that syllables are represented as a “block” that contains the initial consonant(s), defined as the onset, and a “block” that contains both the vowel and word-final consonant(s), defined as the rime, with rimes being

behaviourally relevant for developing literacy performance (e.g., Goswami & Bryant, 1990; Kirtley, Bryant, MacLean, & Bradley, 1989; MacKay, 1972; Treiman, 1983, 1985). For example, it has been shown that reading using rime (word-final-unit) analogies (e.g., *pin* on the basis of *win*) emerges earlier in development relative to reading using body (word-initial-unit) analogies (e.g., *pin* on the basis of *pig*) (Goswami, 1986, 1988, 1991; Goswami & Bryant, 1990). On the other hand, rime advantages do not hold in some other work (Geudens & Sandra, 2003; Geudens, Sandra, & Van den Broeck, 2004; Geudens, Sandra, & Martensen, 2005), and may be task dependent (e.g., Duncan, Seymour, & Hill, 1997; Bowey, Vaughan, & Hansen, 1998; Savage, 2001). We add to this work by comparing children's ability to learn constraints from word-initial and word-final units.

### 1.1. The current study

We assessed 7-year-olds' ability to learn novel graphotactic restrictions either in word-initial units (i.e., between word-initial consonants ( $C_{1s}$ ) and the adjacent following vowel) or in word-final units (i.e., between word-final consonants ( $C_{2s}$ ) and the adjacent preceding vowel). English-speaking (Exp.1) and Turkish-speaking (Exp.2) children were tested using adapted orthographic stimuli. This allows us to generalize our findings across children previously exposed to quite different orthographic systems: Turkish has much more regular sound-to-letter correspondences than English (Öney & Durgunoğlu, 1997).

We replicated the methods of Samara and Caravolas (2014), with two modifications. First, given the greater potential difficulty of learning in this experiment (since redundant cues were removed), exposure occurred over two sessions (rather than one). Secondly, instead of a single-letter detection task, which may have attenuated children's ability to learn two-letter restrictions, we asked children to respond to a change in color across the three letters.

We predicted that both English- and Turkish-speaking children would extract the graphotactic regularities exemplified during training both across conditions (hypothesis-1), and in each condition (hypothesis-2), and stronger learning from word-final than word-initial units in both linguistic contexts (hypothesis-3).

## 2. Methods

### 2.1. Participants

Seventy-eight Year 2 English-speaking children (mean age = 7.24 years) and 37 monolingual Turkish Grade 1 children (mean age = 6.73 years) were recruited from primary schools in England and Turkey, respectively<sup>1</sup>. Note that our original sample was 40 English-

---

<sup>1</sup>While we did not systematically conduct standardized tests of literacy ability, we collected reading scores from the WRAT and TOWRE for a subset of our English-speaking participants. These were as follows: WRAT-IV: mean = 118.00,  $SD = 9.79$ ,  $n = 18$ ; TOWRE: mean = 118.79,  $SD = 11.03$ ,  $n = 57$ . These standardized results suggest that the children we have recruited were above typical levels, possibly due to the fact that we used an opt-in recruitment procedure (as is typical in many developmental studies): that is, parents of higher achieving children are more likely to give consent for them to participate in research. As a further check, for those children where we had available data, we looked for correlations between their literacy scores and their performance on our experimental task: none were present (WRAT:  $r = -.28$ ,  $p = .255$ ; TOWRE:  $r = .11$ ,  $p = .419$ ), suggesting our experimental effects were not carried by exceptional readers.

speaking children; an additional 38 participants were recruited in light of some inconclusive Bayes Factor (BF) results<sup>2</sup>. Participants were randomly allocated to the word-initial condition (45 English-speaking children; mean age = 7.14 years; 19 Turkish-speaking children; mean age = 6.71 years) and word-final condition (33 English-speaking children; mean age = 7.37; 18 Turkish-speaking children; mean age = 6.75 years)<sup>3,4</sup>. All-but-four participants completed two sessions on two consecutive days<sup>5</sup>.

## 2.2. Material

Thirty-two C<sub>1</sub>VC<sub>2</sub> pronounceable English letter strings (30 nonwords, e.g., *gop*; 2 words) were created using four consonant graphemes as C<sub>1</sub>s (*d, g, l, m*), four consonant graphemes as C<sub>2</sub>s (*b, p, r, s*), and *o* and *e* as word-medial vowels. All graphemes and the resulting bigrams were both permissible and frequent within English words in their respective positions. Thirty-two pronounceable Turkish nonwords (e.g., *küç*) were similarly created using different letters from the Turkish alphabet to minimize the presence of unnatural letter strings. In each case, stimuli were arranged into four lists, three of which served as exposure, legal unseen and illegal materials for each participant. Item assignment to list was counterbalanced across participants, such that, stimuli that served as legal items for half of the children were illegal items for the other half, and vice versa.

As shown in Figure 1, for stimuli in the *word-initial condition*, two of the four C<sub>1</sub>s preceded *o* and the remaining 2 C<sub>1</sub>s preceded *e* (e.g., in one counterbalanced list,  $p(d/g, o) = p(l/m, e) = .25$ ) whereas C<sub>2</sub>s followed both *o* and *e* with equal probability ( $p(o, b) = p(e, b) = .125$ ). That is, C<sub>1</sub>s were the only predictive cue of the *adjacent following* vowel's identity. For stimuli in the *word-final condition*, two of the four C<sub>2</sub>s followed *o* and the remaining 2 C<sub>2</sub>s followed *e* (e.g., in one counterbalanced list,  $p(o, b/p) = p(e, r/s) = .25$ ), whereas C<sub>1</sub>s preceded both vowels with equal probability ( $p(d, o) = p(d, e) = .125$ ). That is, C<sub>2</sub>s were the only predictive cue of the adjacent preceding vowel's identity.

Eight pattern-conforming stimuli were presented during exposure and another eight served as legal unseen test items. Eight illegal items (presented at test) violated the patterns.

---

<sup>2</sup>In contrast to the interpretation of *p* values in frequentist analyses, Bayes Factors remain a valid measure of evidence even with optional stopping (Dienes, 2016; Rouder, 2014).

<sup>3</sup>Of the 78 English-speaking children, 69 were monolingual English speakers. The remaining children were reported to be bilingual but were not literate in their second language.

<sup>4</sup>Due to different policies regarding age of school entry in England and Turkey, Turkish-speaking children were significantly younger relative to their English-speaking counterparts,  $t(113) = 5.61, p < .001, d = 1.02$ .

<sup>5</sup>Four Turkish-speaking children completed the sessions over 3 to 6 days.

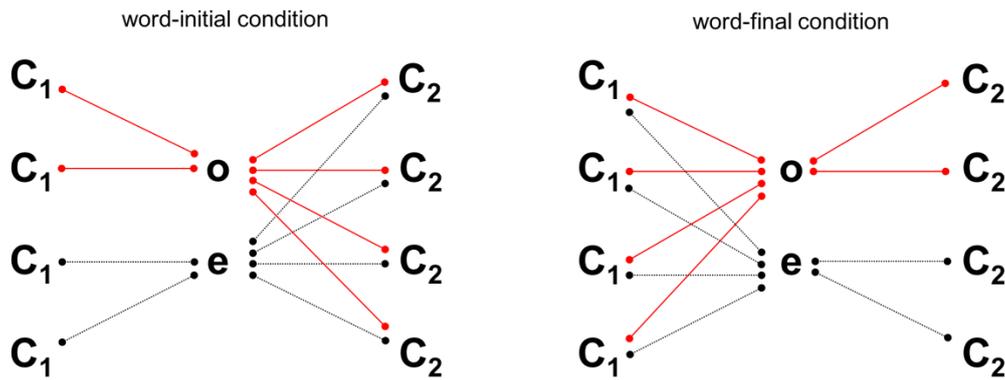


Figure 1. Schematic representation of the underlying graphotactic restrictions in the word-initial and word-final experimental conditions

### 2.3. Procedure

Children were introduced to a toy “froggy” and were invited to play games in his language. The 2-session experimental procedure (controlled using PsychoPy 1.82.01; Peirce, 2007) involved a practice (beginning of session 1), an exposure (spread over the sessions), and a test phase (end of session 2). The practice task involved seeing nine 3-letter English words, printed with black font, and pressing a corresponding key when the stimulus changed color (350ms from stimulus onset). There was no response time limit, and each stimulus was followed by a 500 ms interval. The same procedure was repeated during exposure without any feedback. Six blocks of 48 trials (6 repetitions/string per block) were presented over the two sessions (288 trials; 144/session). At test, children were told that they would see new words, and they had to decide whether they “went well” with “froggy’s language” by pressing on a computer key. They were encouraged to take their time and trust their “gut feeling”. Sixteen test strings (8 permissible/8 impermissible), each followed by a 500 ms interval, were presented in one block without feedback.

### 3. Results

Data and analyses are available at [https://osf.io/vwz8n/?view\\_only=1ffac8a65cc74fc9915b8cb493e8b61c](https://osf.io/vwz8n/?view_only=1ffac8a65cc74fc9915b8cb493e8b61c). We subjected binary response data (i.e., whether an item was endorsed as legal or not) to logistic mixed effects (lme) models, using the lme4 package (Bates, Maechler, Bolker, & Walker, 2015) in R (R Core Team, 2014). Legality and Condition (fixed-effect predictor variables) were coded as centered numerical predictors (so that the intercept represents the grand mean). Random intercepts for subjects and the by-subject random slope for legality were included as random effects<sup>6</sup>. We explored three hypothesis by inspecting fixed-effect model coefficients for the following main effects/interactions: (i) children would discriminate between legal and illegal items across conditions (main effect of legality across conditions), (ii) children would discriminate between legal and illegal items in each condition (main effect of legality in each

<sup>6</sup>We did not include random effects for items since power on this dimension was low, and it is not common for these to be included for work in this area. Including intercepts for items as random effects did not show significant improvements in terms of model fit over the models reported here.

condition), and (iii) learning from word-final context would be greater than learning from word-initial context (legality by condition interaction).

For each coefficient relating to each hypothesis, we computed a Bayes Factor (Bayes Factor) (following Dienes, 2008, 2014) to compare the strength of evidence for  $H_1$  over  $H_0$ . This requires (a) a model of the data: here we used *SEs* and betas for the relevant coefficients (in log-odds space to meet normality assumptions) (b) a model of  $H_1$ : here, in each case, we used a half-normal distribution with an *SD* of  $x$  (Dienes, 2014).  $x$  was determined as follows: *Exp.1 English-speaking children*: for hypothesis (i) and (ii),  $x$  was set to be the learning effect from Samara and Caravolas (2014) (0.19; see Appendix A), making this a rough estimate of the expected effect; for hypothesis (iii) we estimated that a rough maximum effect would be a difference score capturing learning equivalent to that found in the word-final condition and chance performance in the word-initial condition. Thus, we set  $x$  to be *half* of this value ( $x$  is the *SD* of the half-normal and a maximum is approximately  $2SD^7$ ). *Exp.2, Turkish-speaking children*: for (i) and (ii) we again specified rough estimates of the expected effect, however, these were informed by the methodologically relevant learning effects obtained in *Exp.1* (i.e. 0.42); for hypothesis (iii), we constrained  $H_1$  by determining a rough maximum effect, calculated as per above difference score.

Our key inferential statistics are Bayes Factors. These were interpreted using Jeffreys (1961) convention that values  $< 0.33$  suggest substantial evidence for  $H_0$ ; values  $> 3$  suggest evidence for  $H_1$ ; and values between 0.33 and 3 suggest inconclusive evidence. We also computed ranges of values over which substantial Bayes Factors hold<sup>8</sup>.  $p$  values are also reported, although for analyses on English speakers, these are not exact due to the sample size increase outlined in section 2.1<sup>9</sup>.

Results are summarized in Table 1. As predicted, the Bayes Factors showed substantial evidence that more legal than illegal items were endorsed across conditions (hypothesis 1) for both English- and Turkish-speaking participants. Similarly, the Bayes Factors showed substantial evidence, for both groups of participants, that more legal than illegal items were endorsed in *each* of the word-initial and word-final conditions separately (hypothesis 2). With regards to our final prediction regarding performance differences between conditions, the Bayes Factors suggested inconclusive evidence, thus,  $H_1$  could be neither accepted nor rejected in either participant group.

---

<sup>7</sup> An alternative model of  $H_1$  where we have a rough maximum would be to use this as a maximum of a uniform distribution; we choose the current approach in order to bias smaller over bigger effects, as expected in experimental research.

<sup>8</sup>following Dienes (personal communication)

<sup>9</sup>In the original sample of English-speaking children ( $n = 40$ ; see section 2.1), the pattern of significance was identical to that reported here, except for a nonsignificant effect of legality in the word-initial condition.

Table 1. Mean endorsement rates for legal and illegal items (and SDs), Bayes Factors for models (i) – (iii), and lme results.

	Bayes Factor analyses			Frequentist analyses		<i>p</i>
	sample means ( <i>SEs</i> ) <sup>c</sup>	rough estimate of H <sub>1</sub>	BF	Endorsement rates		
				Illegal ( <i>SD</i> )	Legal Unseen ( <i>SD</i> )	
English-speaking children						
Hypothesis 1	0.42 (0.14)	0.19	22.59 <sup>e</sup>	.52 (0.17)	.60 (0.17)	.003
Hypothesis 2a <sup>a</sup>	0.39 (0.18)	0.19	4.36 <sup>f</sup>	.53 (0.16)	.61 (0.16)	.031
Hypothesis 2b <sup>b</sup>	0.46 (0.21)	0.19	3.75 <sup>g</sup>	.49 (0.19)	.58 (0.19)	.033
Hypothesis 3	0.07 (0.28)	0.23 <sup>d</sup>	0.88	n/a	n/a	.809
Turkish-speaking children						
Hypothesis 1	0.55 (0.19)	0.42	26.03 <sup>e</sup>	.44 (0.15)	.54 (0.15)	.004
Hypothesis 2a <sup>a</sup>	0.53 (0.27)	0.42	4.13 <sup>h</sup>	.34 (0.14)	.44 (0.14)	.048
Hypothesis 2b <sup>b</sup>	0.57 (0.28)	0.42	4.51 <sup>i</sup>	.54 (0.17)	.65 (0.17)	.041
Hypothesis 3	0.04 (0.39)	0.28 <sup>d</sup>	0.85	n/a	n/a	.913

*SD* = standard deviation; *SEs* = standard errors; n/a = nonapplicable

<sup>a</sup>word-initial condition

<sup>b</sup>word-final condition

<sup>c</sup>beta coefficients and *SEs* from the relevant lme model (in log-odds space).

<sup>d</sup>Given a maximum twice this value

<sup>e</sup>BF > 3 across all possible betas (*bs*)

<sup>f</sup>BF > 3 for 0.13 < *b* < 1.10

<sup>g</sup>BF > 3 for 0.15 < *b* < 1.25

<sup>h</sup>BF > 3 for 0.22 < *b* < 1.03

<sup>i</sup>BF > 3 for 0.15 < *b* < 1.10

#### 4. General Discussion

Written language is subject to statistical-based spelling rules on the possible/probable successions of graphemes. While some are explicitly taught as spelling mnemonics (e.g., “*i* before *e* except after *c*”), many are not (e.g., doubling is less common before “*ic*” spellings relative to “*ick*” spellings; *magic* vs. *gimmick*). How are these untaught patterns learnt? Following Samara and Caravolas (2014), we investigated whether the same domain-general statistical learning device that operates in spoken language (Saffran, Aslin, & Newport, 1996) is used by young children to generalize over novel letter-context spelling restrictions. 7.5-year-old English-speaking children and 6.5-year-old Turkish-speaking children were incidentally exposed to stimuli that embedded restrictions between vowels and their adjacent preceding/following consonants and, after two training sessions, made legality judgements on novel stimuli that were/were not permissible. Of interest was (i) children’s ability to learn that certain letters cannot co-occur, either word initially (e.g., strings cannot begin with *po*), or word finally (e.g., strings cannot end with *ol*) and (ii) performance differences in their ability to learn from word-initial versus word-final units. We found that English- and Turkish-speaking children learnt the patterns from each type of unit; and there was insubstantial evidence to determine whether children did/did not benefit more from exposure to patterns between word-final rather than word-initial units. We discuss each finding in turn.

The key demonstration is that brief incidental exposure to pattern-embedding stimuli induces learning of novel graphotactic restrictions similar to those found in many alphabetic orthographies. Together with Samara and Caravolas (2014), our learning effects are strong evidence against the view that spellers cannot exploit graphotactic cues while their knowledge of sound-letter connections is still imperfect (cf. logographic stage of spelling development; Frith, 1985; see also Marsh, Friedman, Welch, & Desberg, 1980). This was shown among children learning English (where most work has been carried out) but *also* among children learning the consistent Turkish orthography (whose less demanding nature may attenuate the need for statistical pattern extraction). A strength of the current work is that our result replicates across these two populations, despite their quite different linguistic backgrounds and the use of different stimuli in each case. Critically, we provide substantial evidence that incidental graphotactic learning does *not* depend on the presence of redundant cues: children learned constraints from both word-initial and word-final units in isolation. This further establishes the relevance of this learning mechanism for real-world spelling development. Future work into the limits of children’s statistical learning abilities should assess whether single cues are also easily extracted when they are not exemplified in word edges. Previous phonotactic learning studies suggest that learning word-medial regularities is hard (Endress & Mehler, 2010), thus, more cues may be needed to extract patterns exemplified in these less salient stimulus positions.

Turning to the question of whether children can learn better from word-final versus word-initial units, we found *no evidence* for a stronger cohesion between vowels and word-final consonants, as suggested by one popular view of spelling development (Goswami & Bryant, 1990; Kirtley et al., 1989; Treiman, 1989; Treiman & Kessler, 1995). However, Bayes Factor analyses indicated that  $H_1$  could not be conclusively ruled out. This did not change by collapsing data across experiments 1 and 2 (Appendix B). Supplementary analyses (assuming that the error term would reduce in proportion to  $\sqrt{SE}$ ) suggest that it is not possible to establish  $H_0$  (i.e., demonstrate no difference in children’s ability to learn from these units) even with 200 participants. Larger samples are clearly impractical, thus, different methods are needed to address this question.

A potential source of noise in our experiments is the knowledge participants bring to the task from their own orthographies. Importantly, counterbalancing list assignment means that any biases from native language experience *cannot* be responsible for the learning effects we see (because the items which are “legal” for one half of participants are “illegal” for the other half, and vice versa). One approach to address this in future work is to use an artificial orthography (Taylor, Plunkett, & Nation, 2011).

To conclude, our findings add to emerging evidence on the contribution of statistical learning mechanisms to the acquisition of graphotactic restrictions. Future questions include: what type of knowledge is formed during learning?<sup>10</sup> Similar to Samara and Caravolas (2014), our study was not designed to prevent participants from accessing phonology during training (although verbalization was neither encouraged nor necessary). It is, therefore, possible that learning of graphotactic constraints (words cannot begin with *de*) was complimented by children’s ability to extract the correlated phonological constraints (words cannot begin by /de/). Our ongoing work examines whether graphotactic learning can occur in the absence of phonotactic learning using homophone stimuli (e.g., *co* is legal, *ko* is not).

#### Acknowledgements

This research received no specific grant. We thank Aysé Aktas for collecting the data from Turkish-speaking children.

Declarations of interest: none

#### References

- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1-48. doi:10.18637/jss.v067.i01
- Bowey, J. A., Vaughan, L., & Hansen, J. (1998). Beginning readers’ use of orthographic analogies in word reading. *Journal of Experimental Child Psychology*, *68*(2), 108-133. <https://doi.org/10.1006/jecp.1997.2421>
- Cassar, M., & Treiman, R. (1997). The beginnings of orthographic knowledge: Children's knowledge of double letters in words. *Journal of Educational Psychology*, *89*, 631-644. <http://dx.doi.org/10.1037/0022-0663.89.4.631>
- Dienes, Z. (2008). *Understanding Psychology as a Science: An Introduction to Scientific and Statistical Inference*. Basingstoke: Palgrave Macmillan.
- Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, *5*, 1-17. doi: 10.3389/fpsyg.2014.00781
- Dienes, Z. (2016). How Bayes factors change scientific practice. *Journal of Mathematical Psychology*, *72*, 78-89. doi:10.1016/j.jmp.2015.10.003
- Duncan, L. G., Seymour, P. H. K., & Hill, S. (1997). How important are rhyme and analogy in beginning reading? *Cognition*, *63*, 171-208. doi:10.1016/S0010-0277(97)00001-2

---

<sup>10</sup> We interpret children’s ability to discriminate between legal unseen and illegal test items as sensitivity to a pattern of co-occurrence: “certain consonants only begin/end words only when preceded/followed by certain vowels”. However, each of the eight consonants used in our stimuli featured uniquely as *either* an onset *or* a coda, thus, we cannot rule out the less naturalistic possibility that children were (implicitly) interpreting the pattern as one where ‘certain consonants can only begin/end words only when a certain vowel exist within the same word’. Future experiments featuring some overlapping consonantal onsets/codas should tease apart these different interpretations.

- Endress, A. D., & Mehler, J. (2010). Perceptual constraints in phonotactic learning. *Journal of Experimental Psychology: Human Perception and Performance*, 36, 235-250. <http://dx.doi.org/10.1037/a0017164>
- Frith, U. (1985). Beneath the surface of developmental dyslexia. In K. Patterson, M. Coltheart, & J. Marshall (Eds.), *Surface dyslexia: Neuropsychological and cognitive studies of phonological reading* (pp. 301-330). London: Erlbaum.
- Fudge, E. C. (1969). Syllables. *Journal of Linguistics*, 5, 253-286.
- Fudge, E. C. (1987). Branching structure within the syllable. *Journal of Linguistics*, 23, 359-377.
- Geudens, A., & Sandra, D. (2003). Beyond children's implicit phonological knowledge: No support for an onset-rime structure in children's explicit phonological awareness. *Journal of Memory and Language*, 49, 157-182. [https://doi.org/10.1016/S0749-596X\(03\)00036-6](https://doi.org/10.1016/S0749-596X(03)00036-6)
- Geudens, A., Sandra, D., Martensen (2005). Rhyming words and onset-rime constituents: An inquiry into structural breaking points and emergent boundaries in the syllable. *Journal of Experimental Child Psychology*, 92, 366-387. <https://doi.org/10.1016/j.jecp.2005.07.002>
- Geudens, A., Sandra, D., & Van den Broeck, W. (2004). Segmenting two-phoneme syllables: Developmental differences in relation with early reading skills. *Brain and Language*, 90, 338-352. [https://doi.org/10.1016/S0093-934X\(03\)00446-2](https://doi.org/10.1016/S0093-934X(03)00446-2)
- Goswami, U. (1986). Children's use of analogy in learning to read: A developmental study. *Journal of Experimental Child Psychology*, 42, 73-83.
- Goswami, U. (1988). Orthographic analogies and reading development. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, 40A, 239-268.
- Goswami, U. (1991). Learning about spelling sequences in reading: The role of onsets and rimes. *Child Development*, 62, 1110-1123.
- Goswami, U., & Bryant, P. E. (1990). *Phonological skills and learning to read*. Hillsdale, NJ: Erlbaum
- Hayes, H., Treiman, R., & Kessler, B. (2006). Children use vowels to help them spell consonants. *Journal of Experimental Child Psychology*, 94, 27-42. <https://doi.org/10.1016/j.jecp.2005.11.001>
- Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford: Oxford University Press, Clarendon Press.
- Kessler, B., & Treiman, R. (2001). Relationships between sounds and letters in English monosyllables. *Journal of Memory and Language*, 44, 592-617. <https://doi.org/10.1006/jmla.2000.2745>
- Kirtley, C., Bryant, P., MacLean, M., & Bradley, L. (1989). Rhyme, rime, and the onset of reading. *Journal of Experimental Child Psychology*, 48, 224-245. [https://doi.org/10.1016/0022-0965\(89\)90004-0](https://doi.org/10.1016/0022-0965(89)90004-0)
- MacKay, D. G. (1972). The structure of words and syllables: Evidence from errors in speech. *Cognitive Psychology*, 3, 210-227
- Marsh, G., Friedman, M., Welch, V., & Desberg, P. (1980). The development of strategies in spelling. In U. Frith (Ed.), *Cognitive processes in spelling* (pp. 339-353). San Diego, CA: Academic Press.
- Öney, B., & Durgunoglu, A. Y. (1997). Beginning to read in Turkish: A phonologically transparent orthography. *Applied Psycholinguistics*, 18, 1-15. <https://doi.org/10.1017/S014271640000984X>

- Onishi, K. H., Chambers, K. E., & Fisher, C. (2002). Learning phonotactic constraints from brief auditory experience. *Cognition*, 83, B13-23. [https://doi.org/10.1016/S0010-0277\(01\)00165-2](https://doi.org/10.1016/S0010-0277(01)00165-2)
- Pacton, S., Perruchet, P., Fayol, M., & Cleeremans, A. (2001). Implicit learning out of the lab: The case of orthographic regularities. *Journal of Experimental Psychology: General*, 130, 401-426. doi: 10.1037/0096-3445.130.3.401
- Pacton, S., Sobaco, A., Fayol, M., & Treiman, R. (2013). How does graphotactic knowledge influence children's learning of new spellings? *Frontiers in Psychology*, 4, 1-10. <https://doi.org/10.3389/fpsyg.2013.00701>
- Pearce, J. W. (2007). PsychoPy - Psychophysics software in Python. *Journal of Neuroscience Methods*, 162, 8-13.
- R Core Team (2014). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Rouder, J. N. (2014). Optional stopping: No problem for Bayesians. *Psychonomic Bulletin & Review*, 21, 301-308. doi: 10.3758/s13423-014-0595-4
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274, 1926. doi: 10.1126/science.274.5294.1926
- Samara, A., & Caravolas, M. (2014). Statistical learning of novel graphotactic constraints in children and adults. *Journal of Experimental Child Psychology*, 121, 137-155. doi: 10.1016/j.jecp.2013.11.009
- Savage, R. (2001). A re-evaluation of the evidence for orthographic analogies: A reply to Goswami (1999). *Journal of Research in Reading*, 24, 1-18. doi:10.1111/1467-9817.00129
- Selkirk, E. O. (1982). The syllable. In: H. van der Hulst, & N. Smith (Eds), *The structure of phonological representations* (Part II pp. 337-383). Dordrecht: Foris.
- Taylor, J. S. H., Plunkett, K., & Nation, K. (2011). The influence of consistency, frequency, and semantics on learning to read: An artificial orthography paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37, 60-76. <http://dx.doi.org/10.1037/a0020126>
- Treiman, R. (1983). The structure of spoken syllables: Evidence from novel word games. *Cognition*, 15, 49-74.
- Treiman, R. (1985). Onsets and rimes as units of spoken syllables: Evidence from children. *Journal of Experimental Child Psychology*, 39, 161-181. [https://doi.org/10.1016/0022-0965\(85\)90034-7](https://doi.org/10.1016/0022-0965(85)90034-7)
- Treiman, R. (1986). The division between onsets and rimes in English syllables. *Journal of Memory and Language*, 25, 476-491.
- Treiman, R. (1989). The internal structure of the syllable. In G. Carlson & M. Tanenhaus (Eds.), *Linguistic structure in language processing* (pp. 27-52). Dordrecht, The Netherlands: Kluwer.
- Treiman, R. (1993). *Beginning to spell: A study of first-grade children*. New York: Oxford University Press.
- Treiman, R., & Kessler, B. (1995). In defense of an onset rhyme syllable structure for English. *Language and Speech*, 38, 127-142.
- Treiman, R., & Kessler, B. (2006). Spelling as statistical learning: Using consonantal context to spell vowels. *Journal of Educational Psychology*, 98, 642-652. doi: 10.1037/0022-0663.98.3.642
- Treiman, R., Mullennix, J., Bijeljac-Babic, R., & Richmond-Welty, E. D. (1995). The special role of rimes in the description, use, and acquisition of English orthography. *Journal of Memory and Language*, 124, 107-136.

## Appendix A

### **Linear mixed effect model analyses of children's performance in the contextual constraints learning condition (collapsed across the short and long exposure conditions; $n = 65$ ) in Samara & Caravolas (2014)**

The proportion of items endorsed as legal by child participants in the contextual constraints condition (collapsed across the short and long exposure conditions;  $n = 65$ ) was subjected to a logistic mixed effects model predicting the likelihood of an item being endorsed with legality (legal items, illegal items) as a fixed effect. The model showed a significant effect of legality ( $b = 0.19$ ,  $SE = 0.10$ ,  $z = 1.97$ ,  $p = .049$ ), such that that more legal items ( $M = .46$ ,  $SD = 0.50$ ) were endorsed than illegal items ( $M = .42$ ,  $SD = 0.49$ ). There was also a significant intercept ( $b = -0.27$ ,  $SE = 0.09$ ,  $z = -3.09$ ,  $p = .002$ ), suggesting that children's tendency to reject items (mean endorsement rate = .44,  $SD = 0.50$ ) was reliable.

## Appendix B

### Descriptive statistics, frequentist results and corresponding Bayes Factors across Experiments 1 and 2.

We collapsed across the datasets reported in Experiments 1 and 2 to explore the same three key predictions, i.e., that children would use the underlying statistics to reliably discriminate between legal and illegal items across and in each learning condition (hypothesis 1 and 2), and that learning from word-final units would be greater than learning from word-initial units (hypothesis 3). As shown in Table B2, the Bayes Factors indicated that there was substantial evidence against  $H_0$  for hypothesis 1, 2a, and 2b. The Bayes Factors suggested that  $H_1$  could be neither accepted nor rejected for hypothesis 3.

Table B2. Mean endorsement rates for legal and illegal items (and SDs), Bayes Factors for models (i) – (iii), and lme results across experiments.

	Bayes Factors analyses			Frequentist analyses		
	sample means (SEs) <sup>c</sup>	rough estimate of $H_1$	BF	Endorsement rates		<i>p</i>
				Illegal (SD)	Legal Unseen (SD)	
English-speaking children						
Hypothesis 1	0.46 (0.11)	0.19	636.66 <sup>e</sup>	.49 (0.17)	.58 (0.17)	<.001
Hypothesis 2 <sup>a</sup>	0.42 (0.15)	0.19	16.70 <sup>e</sup>	.48 (0.15)	.56 (0.15)	.004
Hypothesis 2 <sup>b</sup>	0.50 (0.17)	0.19	16.59 <sup>e</sup>	.51 (0.18)	.61 (0.18)	.003
Hypothesis 3	0.08 (0.22)	0.25 <sup>d</sup>	0.83	n/a	n/a	.717

*SD* = standard deviation; *SEs* = standard errors; n/a = nonapplicable

<sup>a</sup>word-initial condition

<sup>b</sup>word-final condition

<sup>c</sup>beta coefficients and *SEs* from the relevant lme model (in log-odds space)

<sup>d</sup>Given a maximum twice this value

<sup>e</sup>BF > 3 across all possible betas