Spelling as statistical learning: evidence from learning experiments with 7 year-old children

Anna Samara

School of Human Sciences, University of Greenwich

CREL, 20/10/2019

Spelling in inconsistent orthographies

/ɛ/

e (bed)

ai (said)

ea (bread)

eo (leopard)

ie

good spelling involves more than pure memorization

Learning to spell in inconsistent orthographies

- Kessler & Treiman (2001): Spelling of inconsistent sound-letter correspondences is actually not that 'chaotic' when probabilistic patterns ('rules') are taken into consideration
- For example:
  - 1) medial /ɛ/ is commonly spelled with an e (e.g., beg) but less frequently before /d/ (e.g., h**ea**d)
  - 3) /ɝ/ is commonly spelled with ur (e.g., curd) but not after /w/ (e.g., w**o**rk, w**o**rth)

Learning to spell in inconsistent orthographies

- Naturalistic (e.g., Treiman, 1993) and some experimental work suggests that children's early spelling attempts reflect 'knowledge' of this sort
  - kat vs *ckat
  - pess vs. *ppes: Children comply to such patterns in pseudoword spelling/2afc tasks (Cassar & Treiman (1997; Pacton et al., 2001)
- Limitations
  - Little control over distributional properties of the input and children's explicit knowledge
  - No insight into the underlying computational mechanisms

Statistical learning

- Basis of humans' **ability to extract statistical patterns** of varying complexity **from the input**
  - e.g., pair frequencies, conditional probabilities btw adjacent elements
- Incidental, 'effortless', unconscious learning
- Key role in language acquisition & development
  - e.g., phonotactics: Infants are sensitive to restrictions on which and where phonemes (or sequences of phonemes) can occur (Jusczyk et al., 1993)
    - English words do not begin with /ŋ/ (but Vietnamese words do)

Statistical learning

- Spoken language research suggests that restrictions on where sounds can occur and which sounds combinations are legal are learnt naturally from early in life
- What about literacy?
  - Literacy acquisition is more protracted
  - Stage models of literacy development (e.g., Frith, 1985; Gentry, 1982)
    - Sensitivity to written language patterns develops at the latest stage of literacy development
  - Testable hypothesis: do statistical learning mechanisms operate in written language from early on?

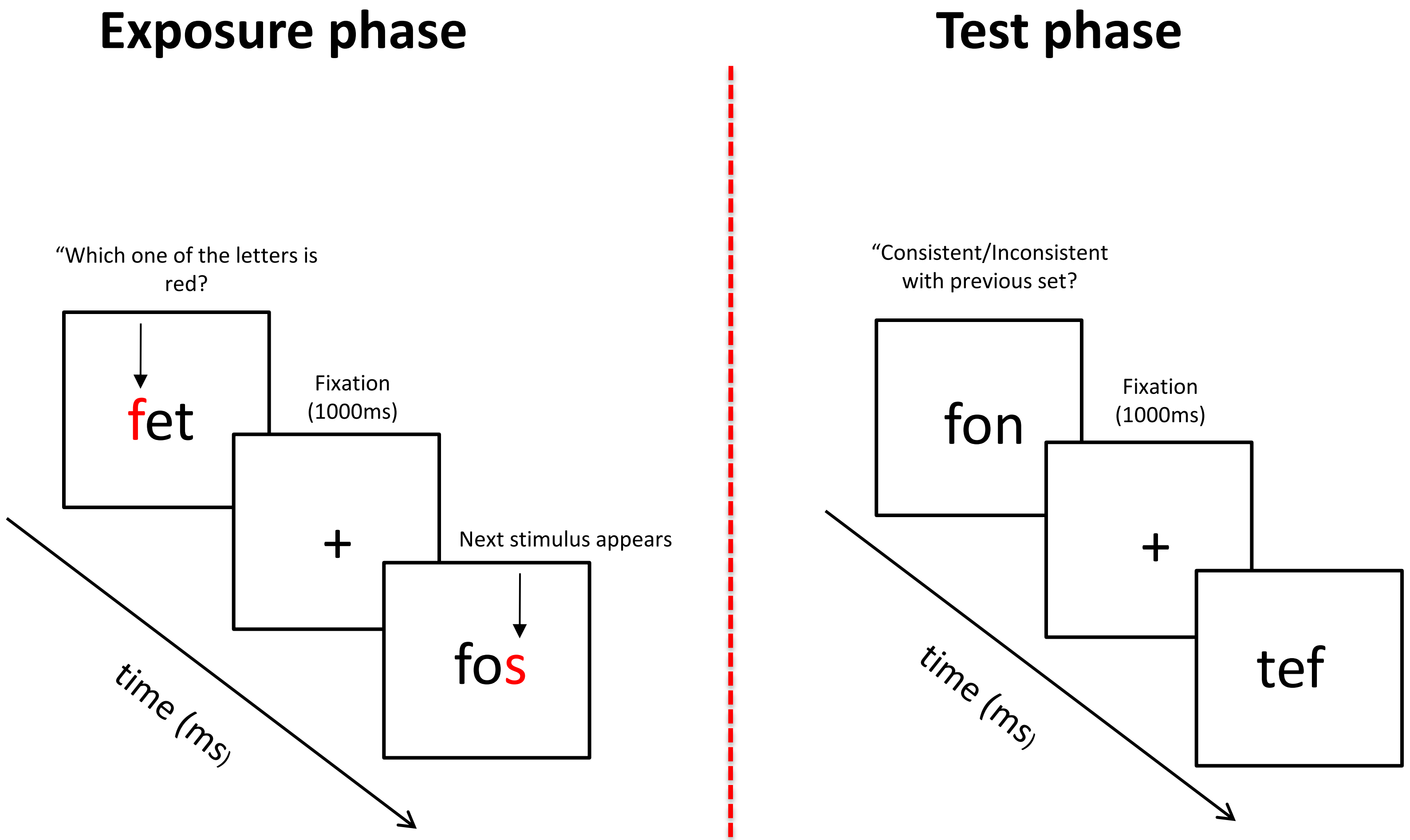
Today: 3 studies

- Exposure to miniature novel lexicon provides full control over input to learning
  - Study 1: Validates these artificial methods in the written language domain
  - Study 2: Addresses further questions regarding orthographic sensitivity in childhood
  - Study 3: Investigates statistical learning of difficult patterns and contrasts it with learning via explicit instruction

Samara & Caravolas (2014). JECP

- Two types of constraints
  - Positional constraints: e.g., words do not begin with ll
  - Context-based patterns: e.g., gz and dz are illegal spellings of frequent word-final sound combinations in English; *bagz, *padz
- Can these be learnt through incidental brief exposure to visual stimuli that embed them?
  - 7 year-olds vs. adult (proficient) spellers

The Incidental Learning task



Stimuli: positional patterns

C | **V** | **C**

d, m, l ,f → o, e → t, n, p, s

d, m, l, f only occur in C_1 position (t, n, p, s cannot)

t, n, p, s only occur in C_2 position (d, m, l, f cannot)

Stimuli: positional patterns

d o t	l o t
d e n	l e n
d o p	l o p
d e s	l e s
m e t	f e t
m o n	f o n
m e p	f e p
m o s	f o s

LU	**IL**
d e t	t o d

Stimuli: context-based patterns

mo ***me*** ***ot*** ***et***

C		**V**		**C**
d, m, l ,f	→	o	→	t, n, p, s
t, n, p, s	→	e	→	d, m, l, f

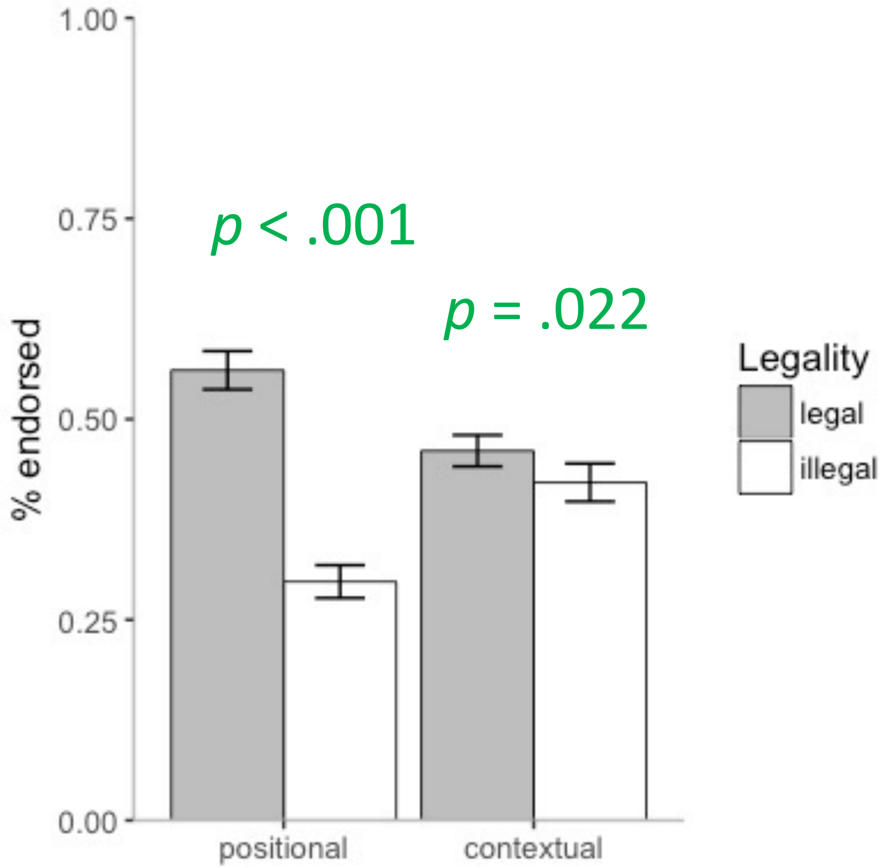
pe ***po*** ***el*** ***ol***

Stimuli: context-based patterns

d o t	t e m
d o p	t e f
m o n	n e d
m o s	n e l
l o t	p e m
l o p	p e f
f o n	s e d
f o s	s e l

LU	IL
d o n	d e t
n e f	l e n

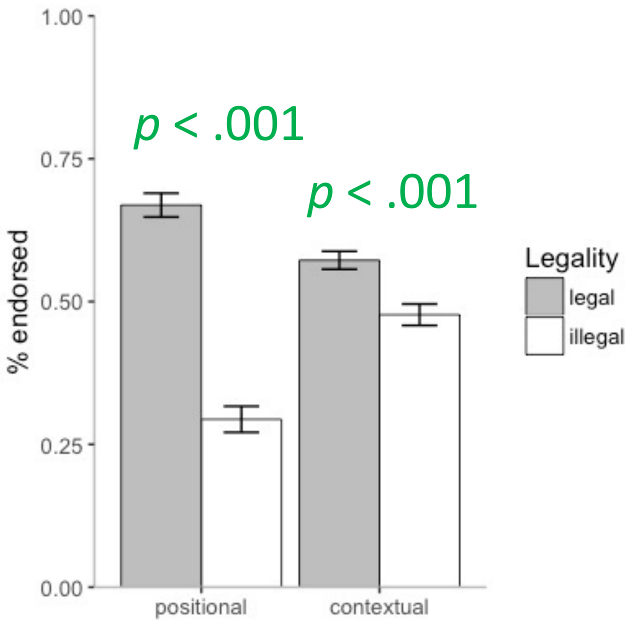
Results: Legality judgments

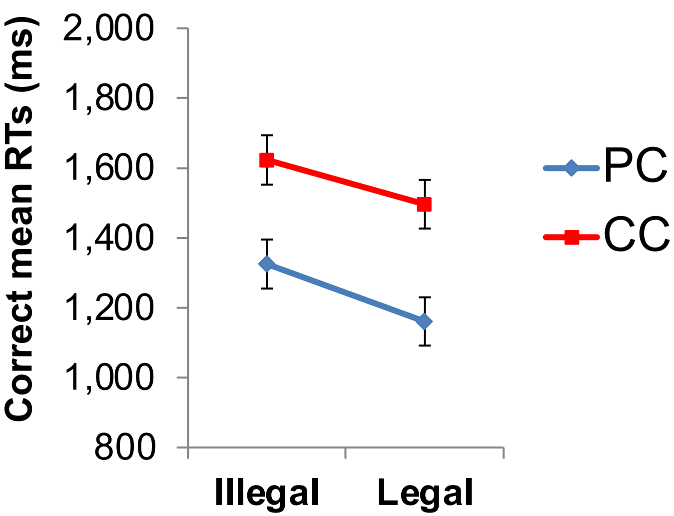


n = 137
mean = 7;5 [years;months]

- Significant learning in both condition
- Learning moderated by pattern complexity (although detection of single letters, e.g., det might have accentuated the difference)

Same pattern of results in adults





Note: Positional patterns learned more reliably than contextual patterns (and adults are, overall, better learners than children)

In sum…

- Study 1 provides evidence that novel positional and context-based patterns can be learn under brief incidental experimental conditions
- Suggests that statistical learning processes operate among 7-year-olds and underlie this ability

Limitations

- Redundancy of cues: Above chance learning under highly favourable conditions…
  - PC learning: constraints on the position of single letters (e.g., ~~*d*~~), as well as bigrams: (CVs; e.g., ~~*de*~~) and rime-level unit (e.g., ~~*et*~~) constraints
  - CC learning: Constraints are exemplified both in word beginnings (e.g., ~~*de*~~) and ends (e.g., ~~*et*~~)
- <u>Are both word contexts necessary for learning to occur?</u>
- <u>If not, are they equally beneficial to learners?</u>

Study 2: Incidental learning of context-based patterns within word-initial (CV) vs. rime-level (VC) units: Evidence from English and Turkish

Samara, Singh, & Wonnacott (2019). Cognition

Study 2: Rationale

- More naturalistic design: Can patterns *in each position* can be learned independently?
- Word-initial (CV) vs. rime-level (VC) comparison
  - Some studies in reading and oral language have shown that rimes (i.e., vowel and word-final consonant(s)) have behavioural relevance for developing and skilled literacy performance
  - If rimes are special, learning patterns from such units should be stronger than learning from CV units

Patterns in word-initial (CV) vs. rime (VC) units

– E.g. 1: medial /ɛ/ is commonly spelled with an e (e.g., beg) but less frequently before /d/ (e.g., head)

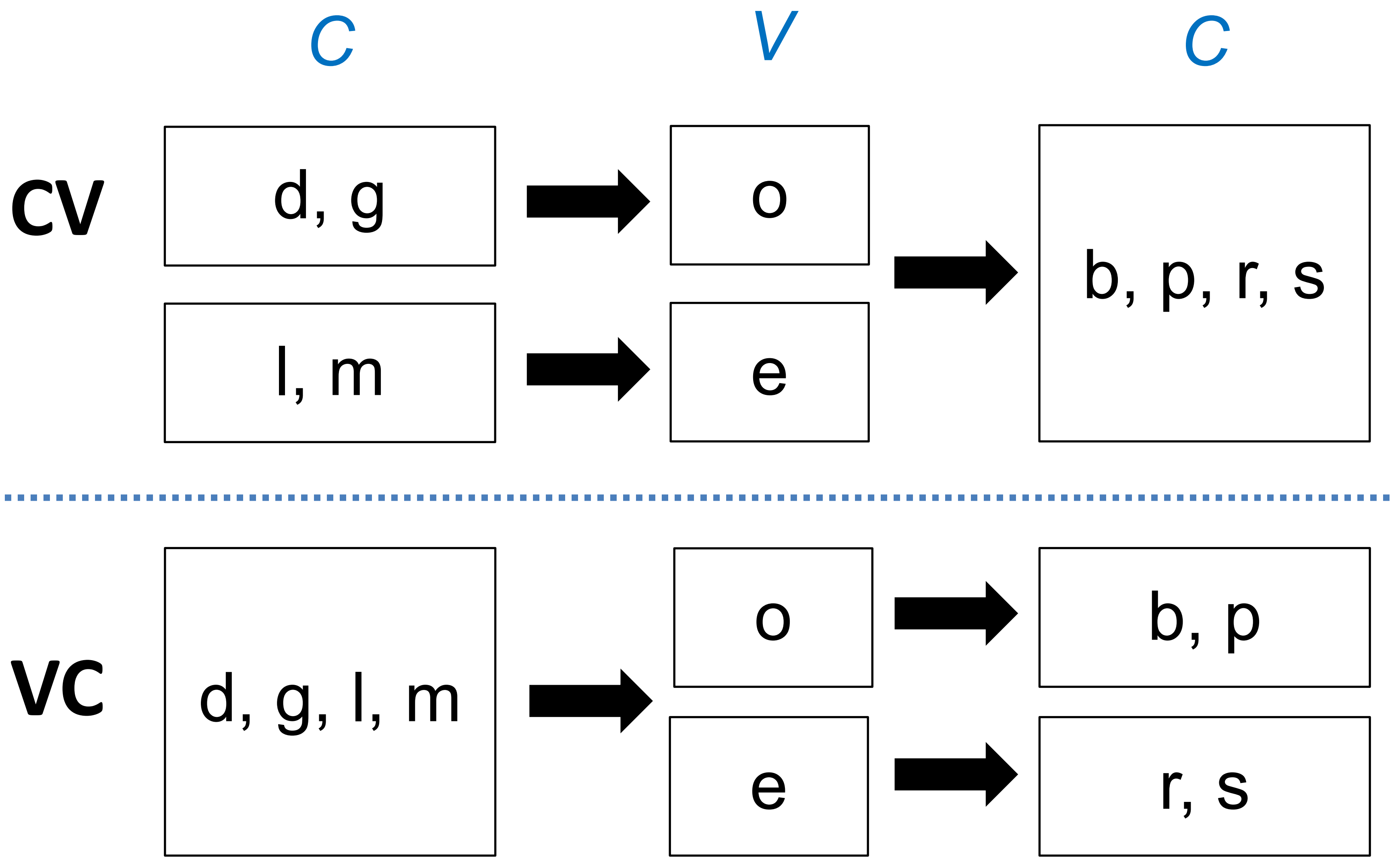
– E.g. 2: /ɝ/ is commonly spelled with ur (e.g., curd) but not after /w/ (e.g., w**o**rk, w**o**rth)

Less numerous & weaker CV contingences in English (Kessler & Treiman, 2001)

Methods & Procedure

- 78 English-speaking children
  - CV condition: n = 45 (mean age = 7.14 years)
  - VC condition: n = 33 (mean age = 7.37 years)
- 37 Turkish-speaking children
  - CV condition: n = 19 (mean age = 6.71 years)
  - VC condition: n = 18 (mean age = 6.75 years)
- Variant of the IGL task introduced in study 1
  - Learning spread across 2 days; tested on day 2
  - Exposure cover task: respond to the stimulus color

Stimuli



Data analyses

- Bayes Factor (BF) analyses
  - Based on the principle that evidence supports the theory that most strongly predicts it
  - Likelihood ratio that indicates the relative strength of evidence for **two theories/models**
  - H_0 vs. H_1
  - **Allows for three type of conclusions:**
    - > 3: Substantial evidence for the alternative
    - < 1/3: Substantial evidence for the null
    - 1/3 < BF < 3: Data insensitivity

Data analyses

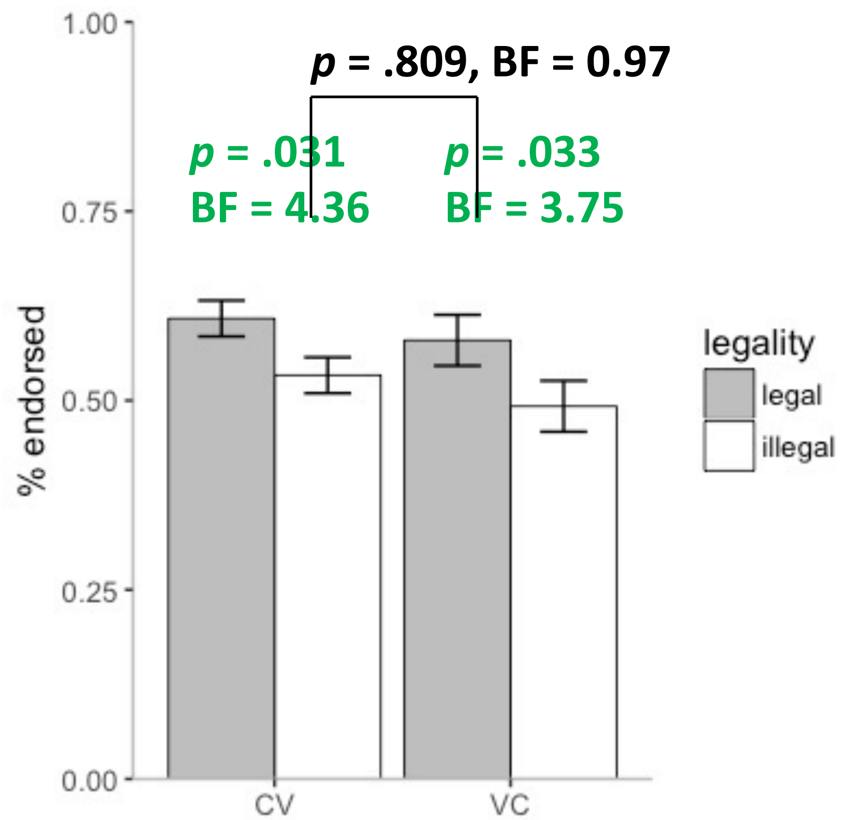
- **Model1:** Model predicting above chance learning performance *(vs. model predicting chance performance in in CV condition)*
- **Model2:** Model predicting above chance learning performance (*vs. model predicting chance performance in VC condition)*
- **Model3:** Model predicting a performance advantage in the word-final (VC) relative to the word-initial (CV) condition (vs. model predicting CV = VC performance)

Priors for H_1 models

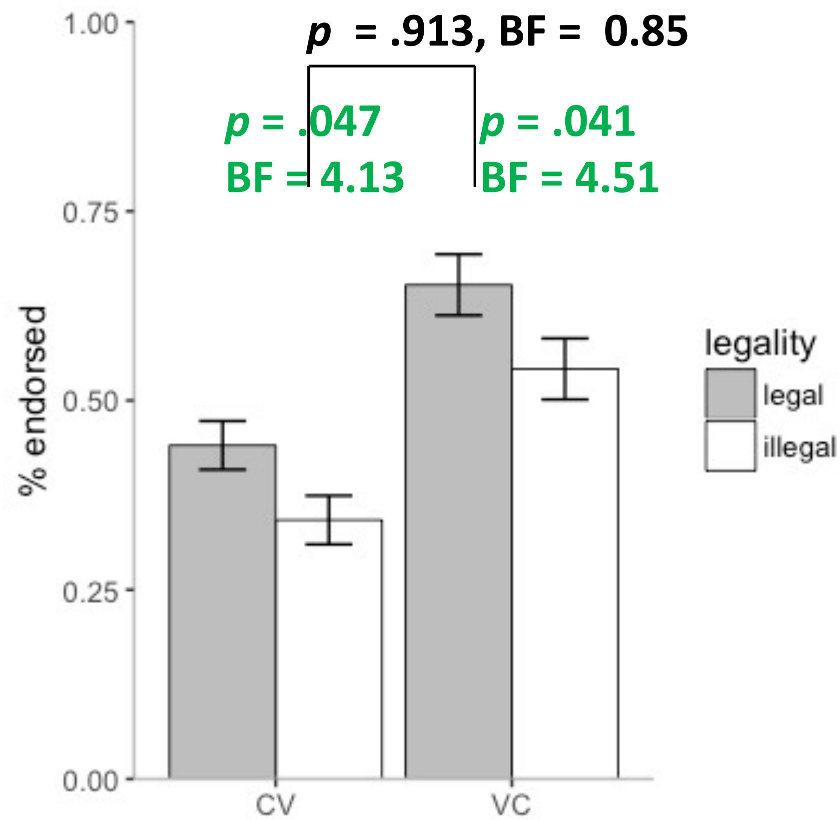
- **Model1:**
  - **H1:** Predicted ES = learning equivalent to that reported for contextual learning in study 1
- **Model2:**
  - **H1:** Predicted ES = learning equivalent to that reported for contextual learning in study 1
- **Model3:**
  - H1: Rough maximum predicted ES (given lack of comparable data) driven from the experiment per se

Results

English-speaking



Turkish-speaking



In sum…

- Substantial learning of novel context-based patterns both within CV (body) and VC (rime-level) units
- Findings replicate in two linguistic contexts (and hold when we collapse across the 2 datasets)

In sum…

- No evidence of the predicted word-final advantage
  – Power analyses show that, based on our current level of variance, approximately 1000 (!) participants are needed to provide evidence of no difference in performance between conditions (i.e., if the true mean difference between conditions was actually zero)

Incidental learning of written patterns with no phonological counterpart

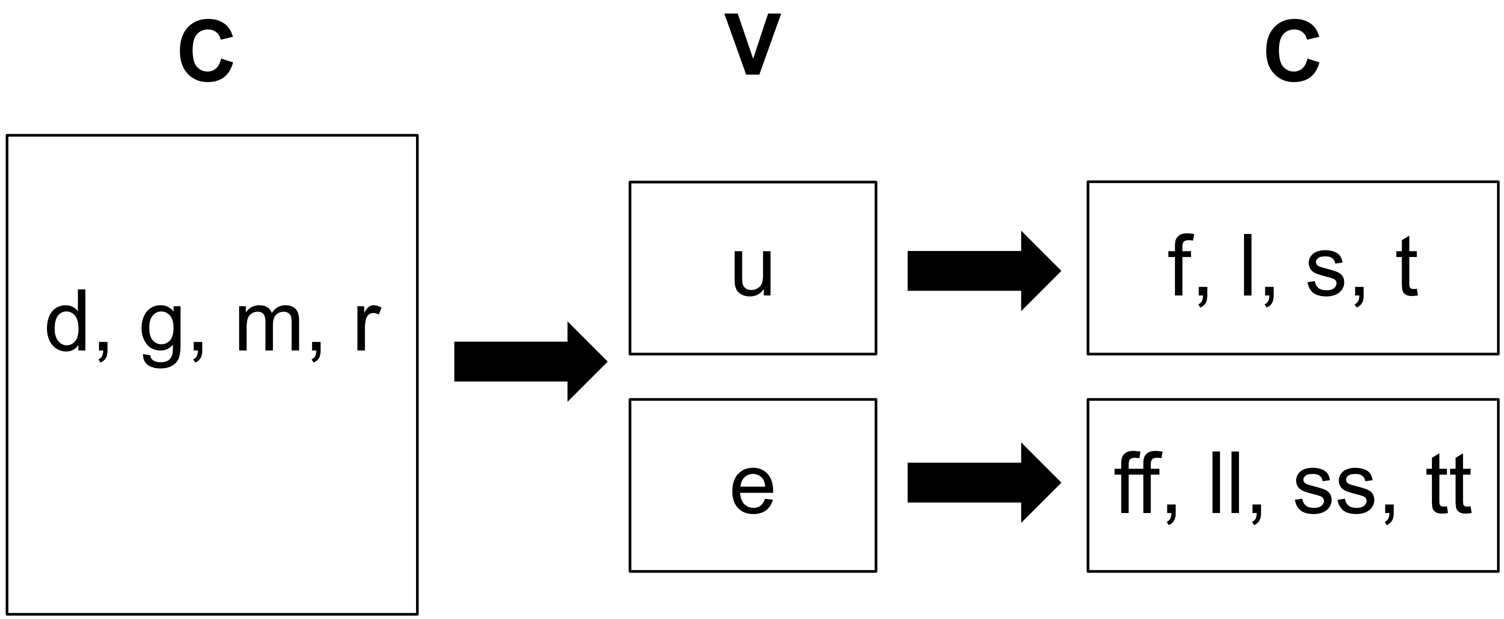
- Spoken/written English words do not begin with *ng
- Purely visual orthographic ‘rules’ that place **constraints** on **where and when** certain **letters** (or letter combinations) can **occur**
- Some of them may be easy to verbalize and may be explicitly taught
  - e.g., gz and dz are illegal spellings of frequent word-final sound combinations in English; *bagz, *padz)
- But others are not….
  - Letters double more often after single-letter-vowel spellings than double-letter-vowel spellings: bedding vs. heading, Jeff vs. deaf etc

Samara, Singh, & Wonnacott (in preparation)

Incidental learning of written patterns with no phonological counterpart

e.g., list1

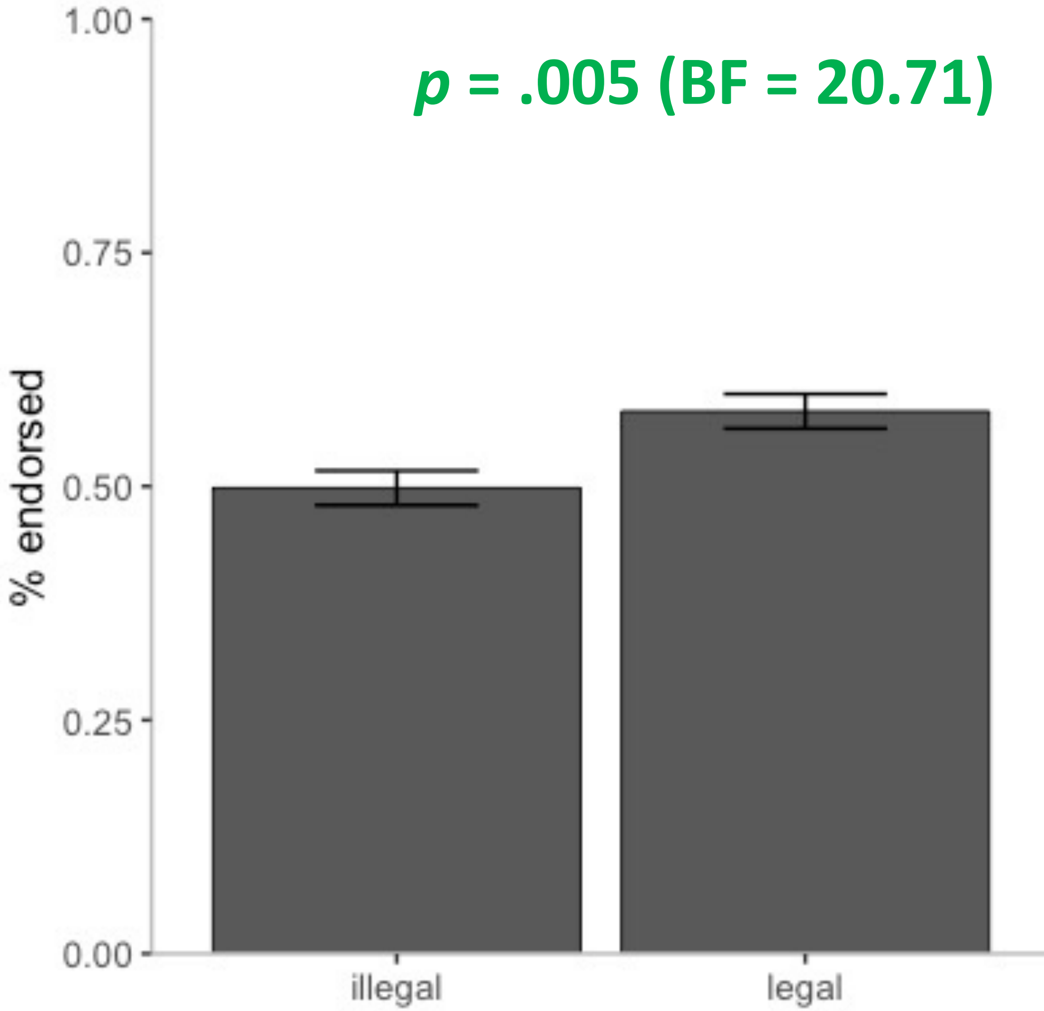
duf rut *duff *rutt



deff rett *det *ret

note: incidental exposure in the context of one-back task

Results: Legality judgments (exp3a)



***p* = .005 (BF = 20.71)**

n = 35
mean = 6.6 years

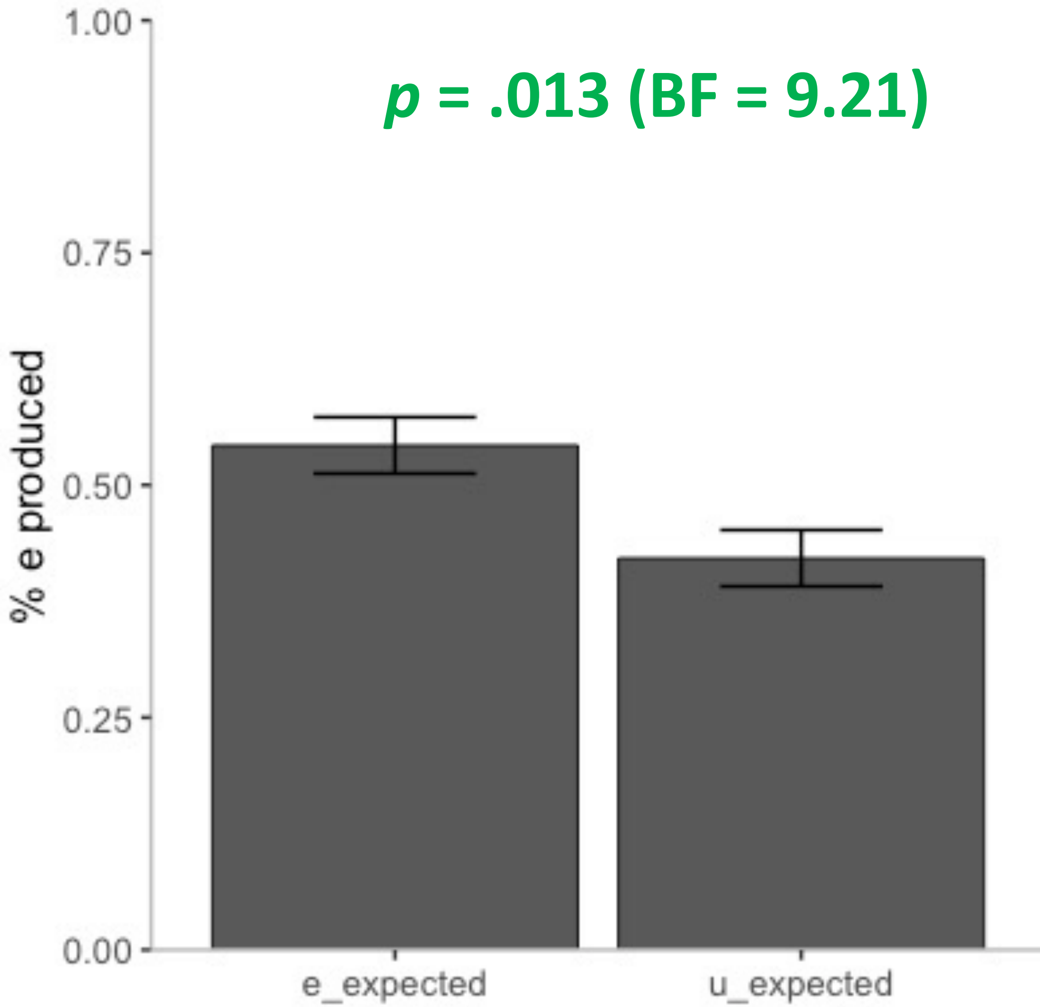
(Predicted learning equivalent to that reported for learning VCs in study 2)

Results: Fill-in-the blanks (exp3a)

d _ ff

e **u**

Results: Fill-in-the blanks (exp3a)



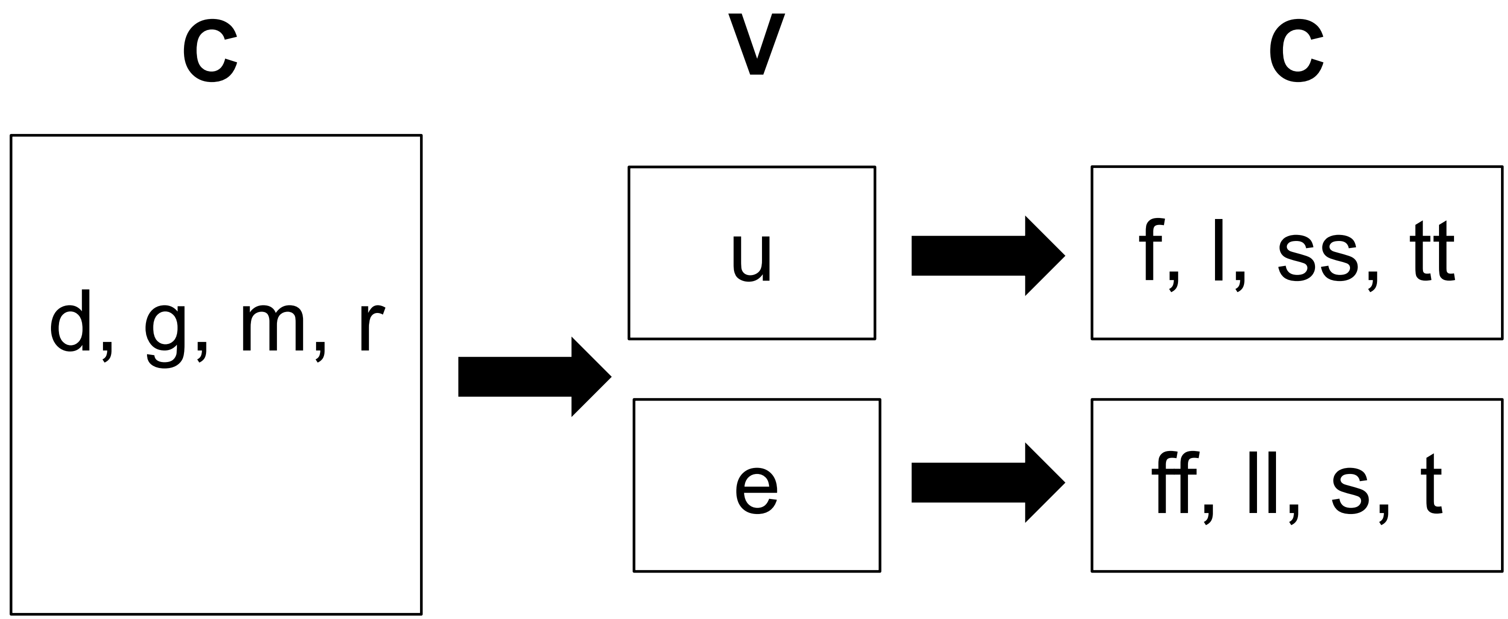
n = 35
mean = 6.6 years

(Predicted learning equivalent to that found in pilot study)

Incidental learning of graphotactics with no phonological counterpart (v2)

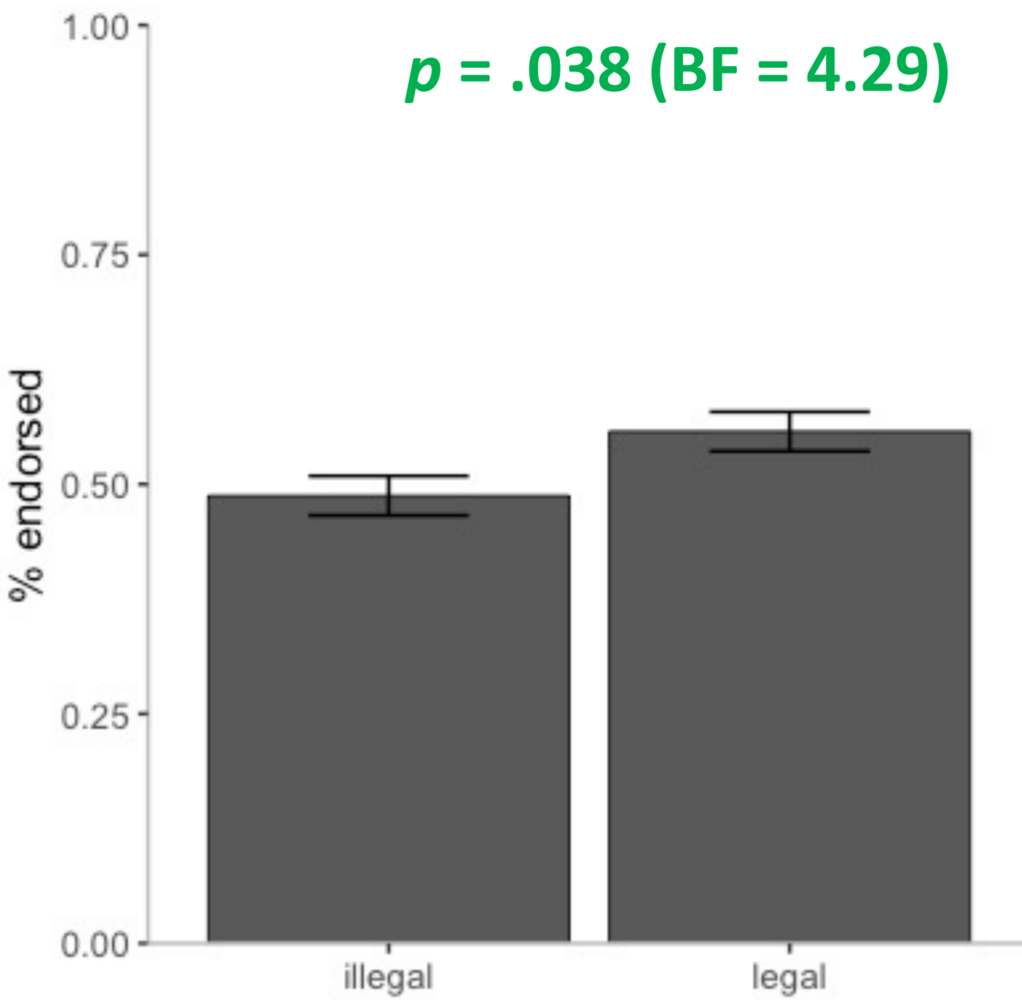
e.g., list1

duf rutt *duff *rut



deff ret *def *rett

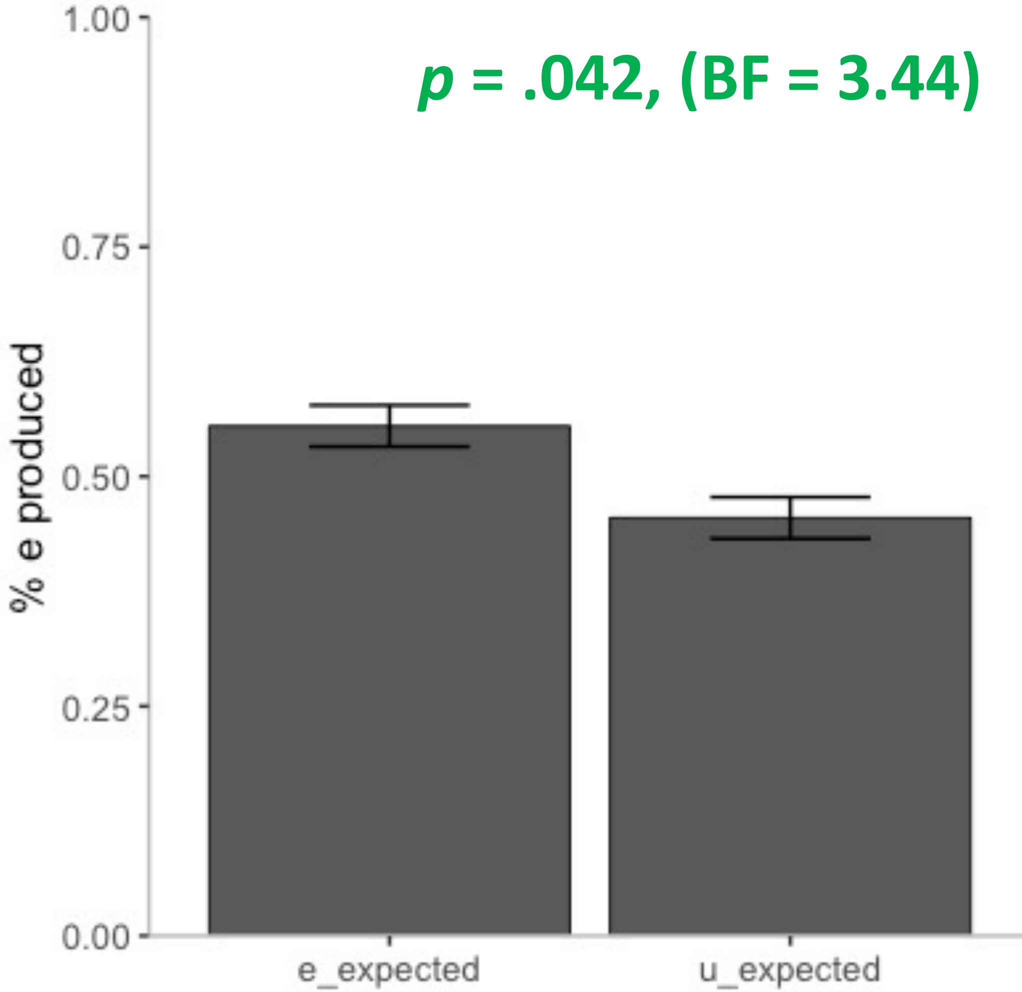
Results: Legality judgments (exp3b)



n = 25
mean = 6.8 years

(Predicted learning equivalent to that reported for learning VCs in study 2)

Results: Fill-in-the blanks (exp3b)



***p* = .042, (BF = 3.44)**

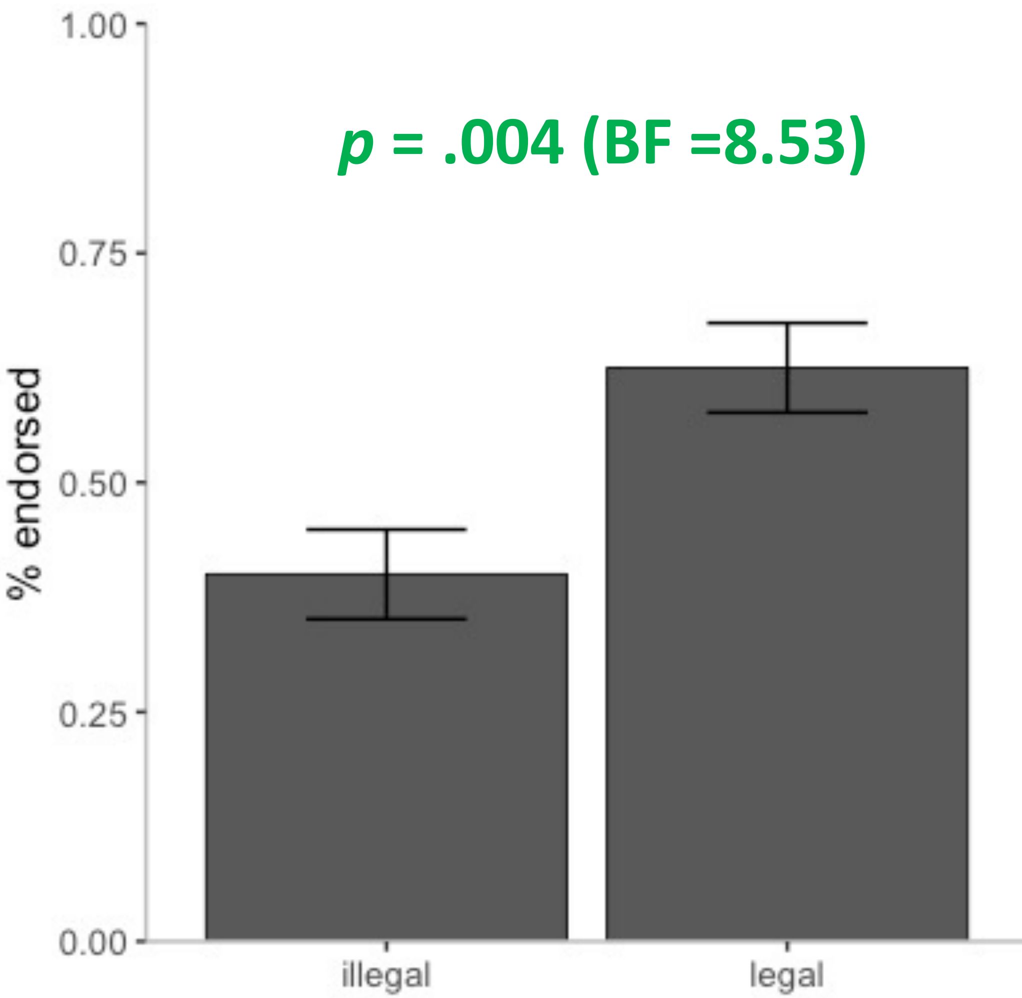
n = 25
mean = 6.8 years

(Predicted learning equivalent to that found in pilot study)

Explicit learning of written patterns with no phonological counterpart

- Children also gain a great deal of knowledge about spelling through explicit instruction
- Should spelling be taught (e.g., Moats, 2005) or caught? (e.g., Wilde, 1990)
- Follow-up of exp.3a
- Patterns explicitly taught: “In Freddie’s language, double letters come after “u” and single letters come after “e” + 2 examples

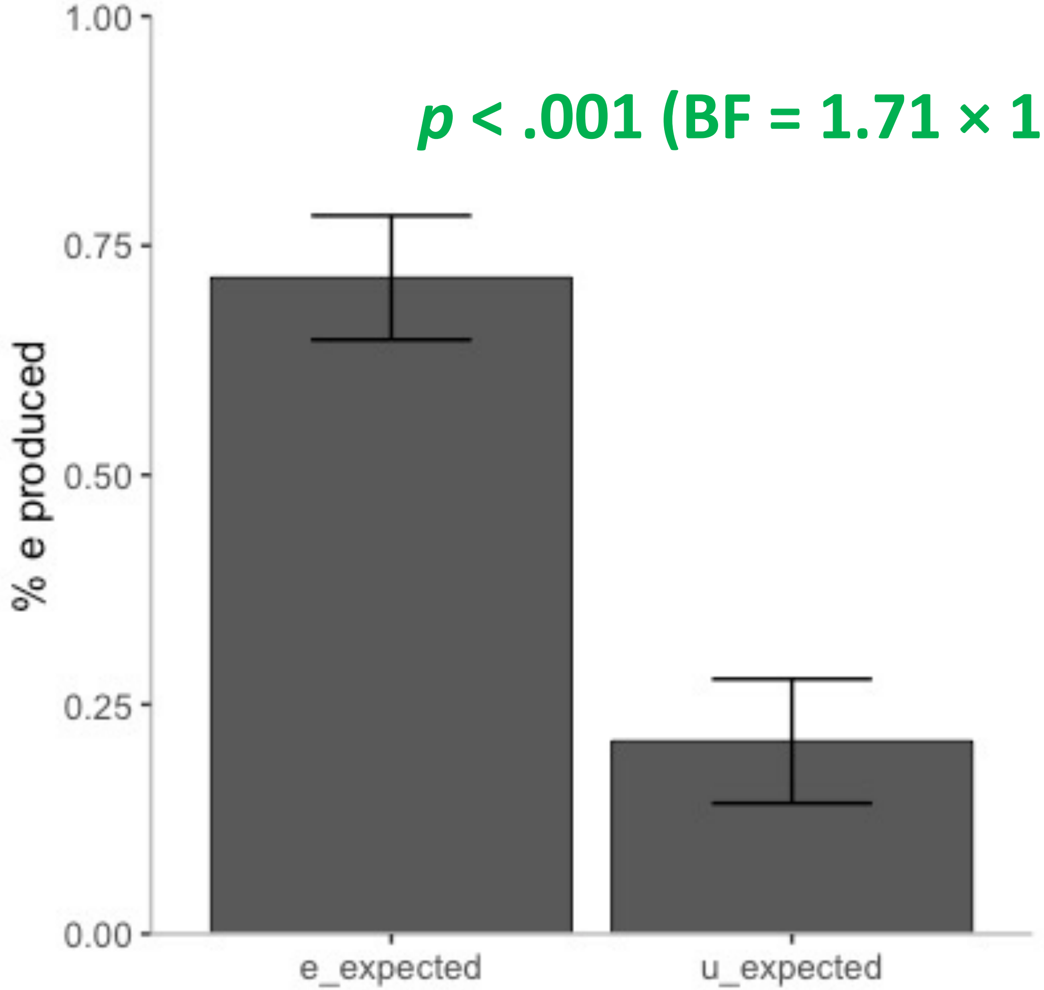
Results: Legality judgments (exp3c)



n = 25
mean = 7.2 years

(Predicted learning equivalent to that reported for learning VCs in study 2)

Results: Fill-in-the-blanks (exp3c)



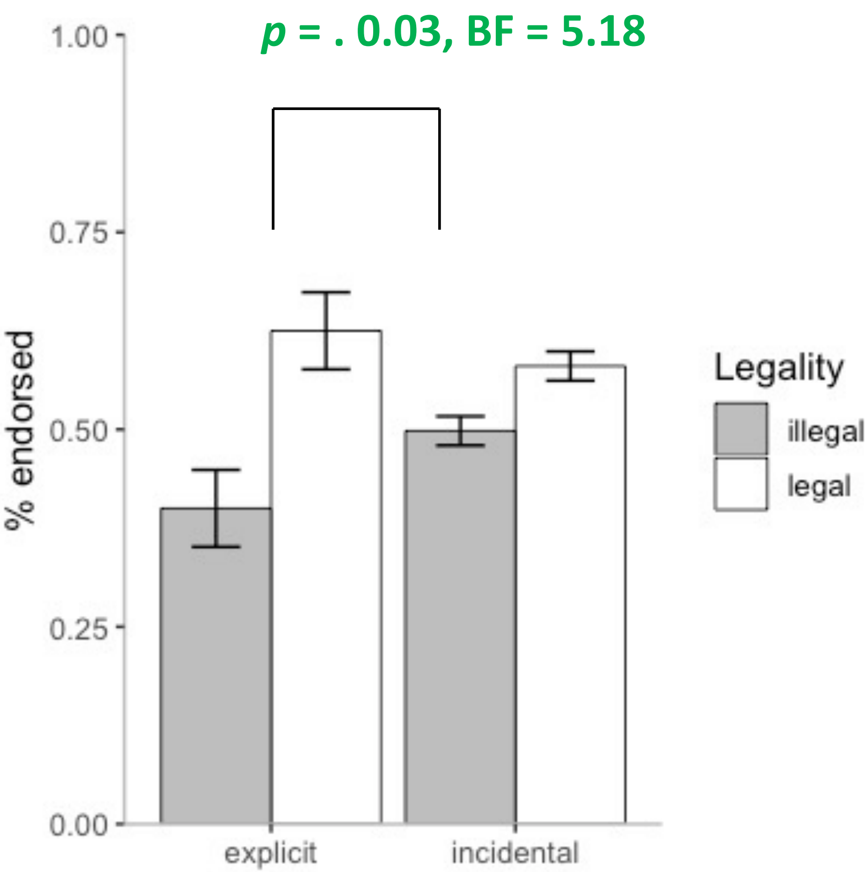
$p < .001$ (BF = 1.71×10^{17})

n = 25
mean = 7.2 years

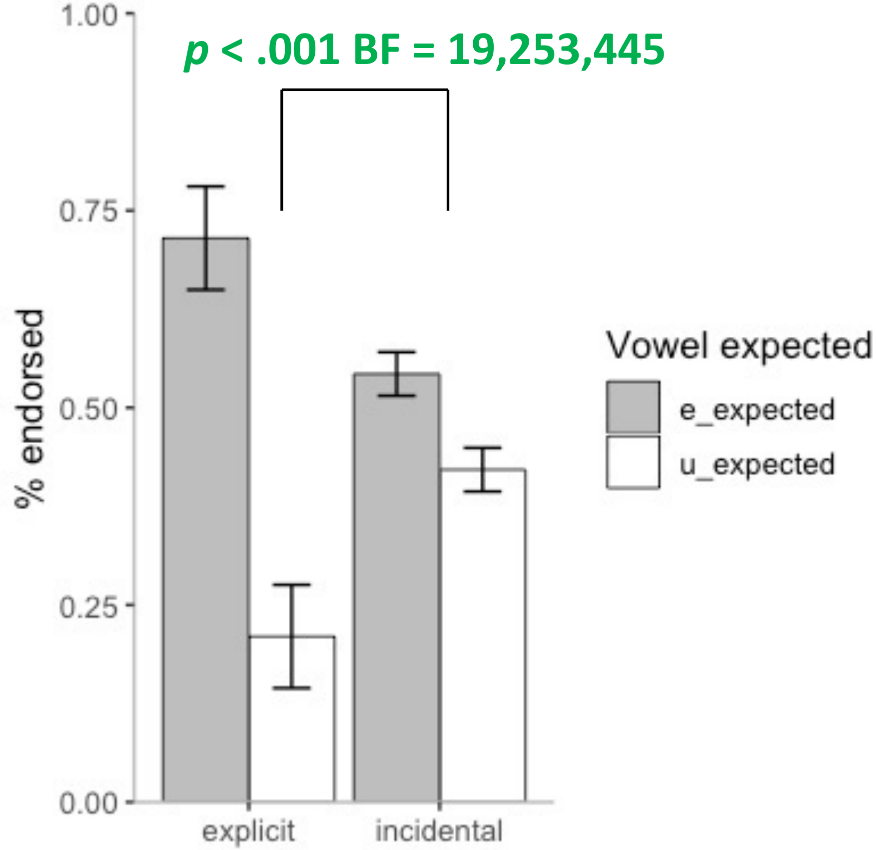
(Predicted learning equivalent to that found in pilot study)

Experiment 3a (incidental) vs. 3c (explicit)

Legality judgments



Fill-in-the-blanks



In sum…

- Substantial evidence that novel context-based patterns within rime-level (VC) are learnt by 6.5 - 7-year-olds when presented under incidental exposure conditions
- Patterns are also readily learnt under explicit training conditions
  - Direct comparison of implicitly vs. explicitly induced learning effects suggests clear advantage of explicit instruction

Bringing it all together

- Study 1
  - Validates methods in written language domain
  - Demonstrates that, from 7 years of age, children are sensitive to novel positional and context-based patterns
- Study 2
  - Employs similar methods to address further questions regarding orthographic sensitivity in childhood
  - Establishes that redundant cues are not necessary for learning to occur
- Study 3
  - Establishes that purely orthographic patterns can be also learned incidentally but benefit from explicit teaching instruction

Bringing it all together

- Implications for theories of literacy development
  - Elucidate the learning mechanisms that allows pattern sensitivity to emerge in the absence of explicit instruction
  - Argues against "late" stage-based models of literacy development (Frith, 1985; Gentry, 1982) by showing that (at least some) orthographic learning occurs early
  - Corroborates a statistical learning account of learning to spell (Pollo et al., 2008; Treiman, 2017; Treiman & Boland, 2017) but also demonstrate the effetiveness of explicit instruction

Directions for future research

- **Exploring the role of statistics**
  - Are children sensitive to conditional forward and backward probability (e.g., the probability that q is followed by u and that u is preceded by q), which may be more relevant for learning to spell than joint probability (e.g., frequency of qu)?
  - Are children sensitive to manipulations of (more naturalistic) probabilistic orthographic patterns? (head vs. bed)
    - Are these exceptions learnt best in a staged manner (i.e., whereby patterns are learnt and consolidated first before exceptions are introduced)?
- **Homophone learning**
  - Does the EL advantage hold across ages and types of patterns
  - How long lasting are implicitly vs. explicitly induced learning effects?

Acknowledgements

Marketa Caravolas *(Bangor)*

Elizabeth Wonnacott *(UCL)*

Daniela Singh *(UCL)*

Acknowledgements

- Marie Curie actions
- The schools
- Research assistants & students
  - Sam Crewe
  - Ayse Aktas

Thank you for listening!

anna.samara@ucl.ac.uk

Adults-study 1

Measure	Range	Mean (*SD*)	1	2	3	4	5	6
		PC learning (*n* = 55)						
1. IGL	-0.16 – 2.75	1.15 (0.70)	—	-.14	.07	.12	.04	-.12
2. WRAT Reading[a]	49.00 – 66.00	60.47 (3.94)		—	.59**	.46**	.01	.33*
3. WRAT Spelling	38.00 – 57.00	47.00 (3.21)			—	.56**	.06	.46**
4. Exception Words[b]	69.00 – 79.00	74.89 (2.42)				—	.32*	.46**
5. TOWRE-SWE[c]	1.49 – 2.68	2.13 (0.22)					—	.50**
6. TOWRE-PDE[c]	0.73 – 1.75	1.28 (0.20)						—
		CC learning (*n* = 56)						
1. IGL	-0.36 – 1.16	0.27 (0.36)	—	-.05	-.16	.13	.18	.01
2. WRAT Reading	51.00 – 67.00	60.23 (3.54)		—	.63**	.72**	.35	.64**
3. WRAT Spelling	37.00 – 55.00	46.57 (3.88)			—	.62**	.13	.48*
4. Exception Words[b,d]	71.00 – 79.00	75.74 (1.91)				—	.38	.65**
5. TOWRE-SWE[c,d]	1.58 – 2.55	2.12 (0.21)					—	.52*
6. TOWRE-PDE[c,d]	0.76 – 1.77	1.26 (0.20)						—

Kids-study 1

Measure	Range	Mean (*SD*)	1	2	3	4
		PC learning (*n* = 60)				
1. IGL	-0.81 – 2.68	0.83 (0.73)	—	.19	.10	.13
2. Reading[a]	6.00 – 119.00	73.55 (22.93)		—	.85**	.77**
3. NW Reading[a]	9.00 – 70.00	35.45 (14.68)			—	.71**
4. PWM[b]	6.00 – 61.00	29.13 (10.25)				—
		CC learning (*n* = 62)				
1. IGL	-0.81 – 1.47	0.15 (0.46)	—	-.11	-.09	-.09
2. Reading[a]	6.00 – 123.00	78.06 (20.77)		—	.83**	.72**
3. NW Reading[a]	2.00 – 72.00	39.94 (18.23)			—	.67**
4. PWM[b]	5.00 – 52.00	30.19 (9.98)				—

Dyslexic adults

Variable	Skilled readers (n = 30)	Dyslexic readers (n = 19)
WRIT Vocabulary	-.15	.32
WRIT Matrices	-.16	.34
WRAT Reading	-.26	.22
WRAT Spelling	-.04	.26
WAIS Digit Span	.15	-.16[b]
WAIS Symbol Search	.04	.44
RAN digits mean time[a]	-.07	.04
RAN objects mean time[a]	.11	.01
NWPD latencies	.15[b]	.01[a]

Note. WRIT = Wide Range Intelligence Test; WRAT = Wide Range Achievement Test; WAIS = Wechsler Adult Intelligence Scale; RAN= Rapid Automatized Naming; NWPD = NonWord Phoneme Deletion.
[a]In seconds. [b]Log transformed.