

Online Experiments for Language Scientists

Lecture 5: Frequency learning

Kenny Smith

kenny.smith@ed.ac.uk

Reminder about Assessment 1

- Due on 10th November
- Read the assignment brief (<https://kennysmithed.github.io/oels2022/assessment/AssignmentBrief2022.pdf>)
- I'll set aside time for questions in next week's lecture
- No questions after 10am on Monday 7th November!

Ferdinand, Kirby & Smith (2019)

Ferdinand, V., Kirby, S., & Smith, K. (2019).
The cognitive roots of regularization in
language. *Cognition*, 184, 53-68.

Large frequency-learning experiment run on
MTurk

- Do domain (linguistic vs non-linguistic) and
demand (tracking 1 vs 6 frequency
distributions) influence **regularization
behaviour?**



Vanessa Ferdinand
(formerly Edinburgh,
now Melbourne)



Simon Kirby
(Edinburgh)

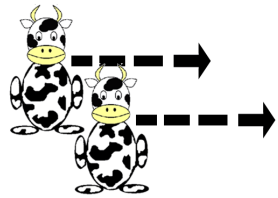
Variation in language

Languages exhibit variation at all levels (paraphrase, synonymy, allomorphy, allophony), but variation is **constrained**

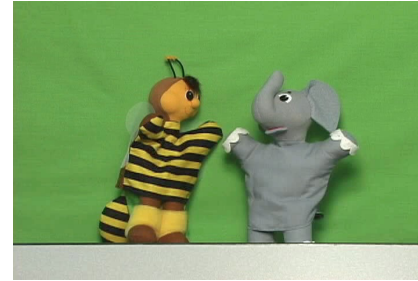
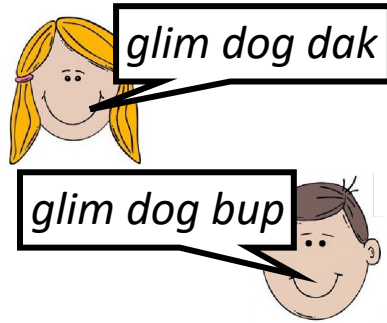
- Languages have lexicons and grammars
- Linguistic (phonological, lexical, syntactic, semantic) or sociolinguistic **conditioning** of alternation
 - English past tense allomorphy: hunt/**ɪd**/ vs fish/**t**/
 - Noun classes: ***la** chaise, **le** sofa, **la** fille, **le** garçon*
 - T-glottaling: glo/**t**/al vs glo/**ʔ**/al

Why is language like this?

Variation-learning experiments



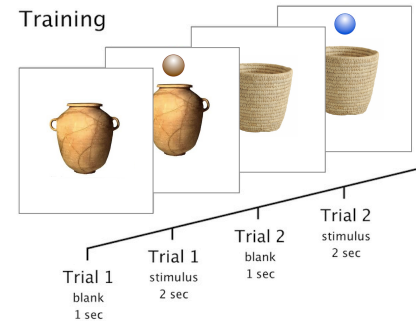
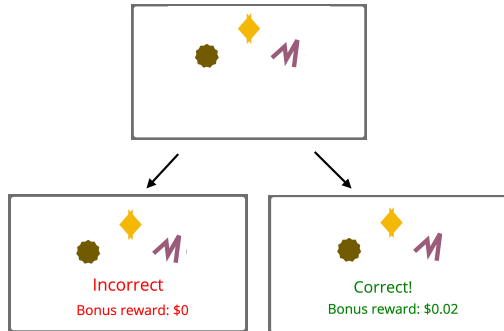
glim cow fip
glim cow tay



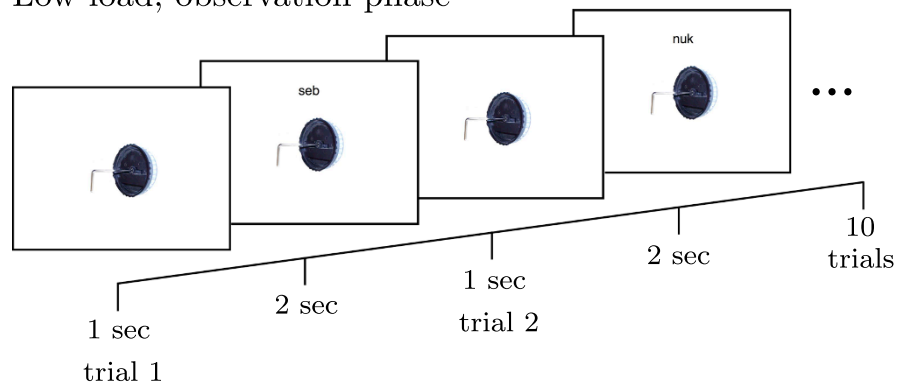
ooshra buzzo trunko
ooshra trunko tid buzzo



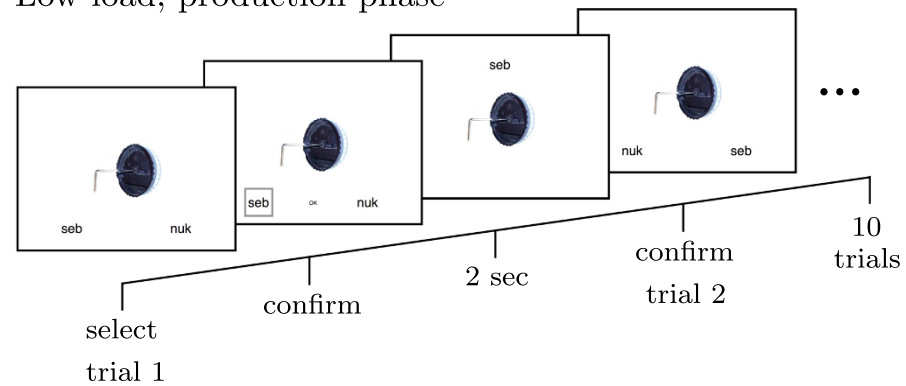
buv
kal



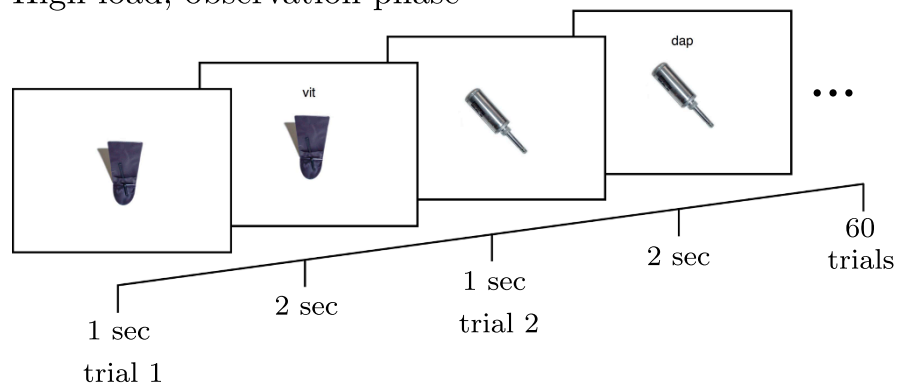
Low load, observation phase



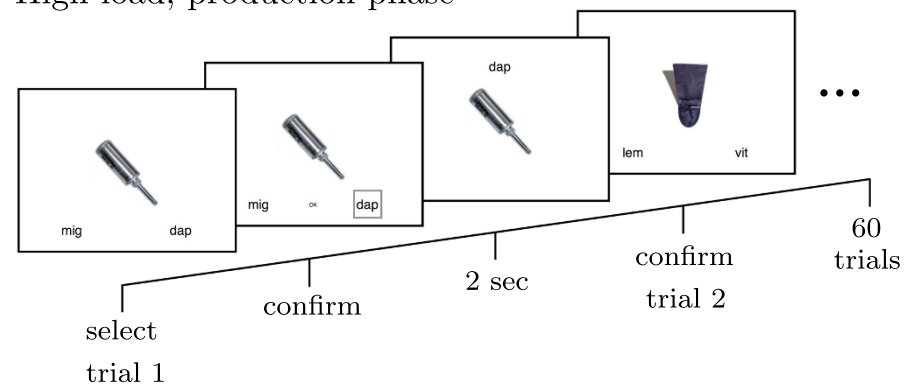
Low load, production phase

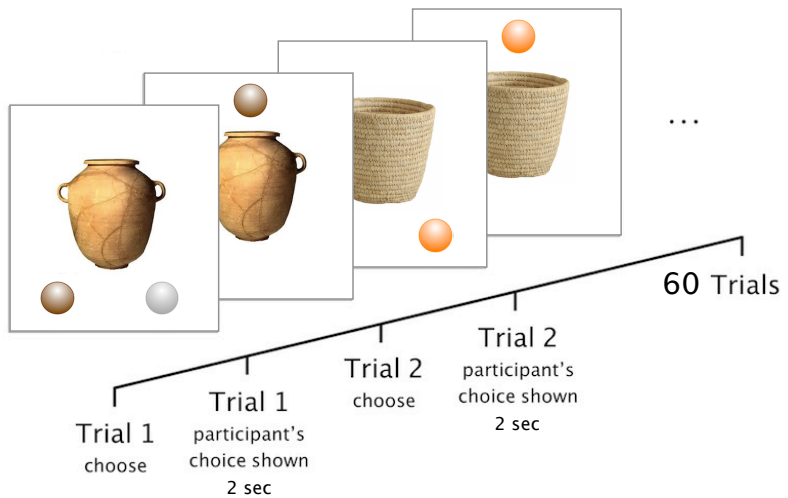
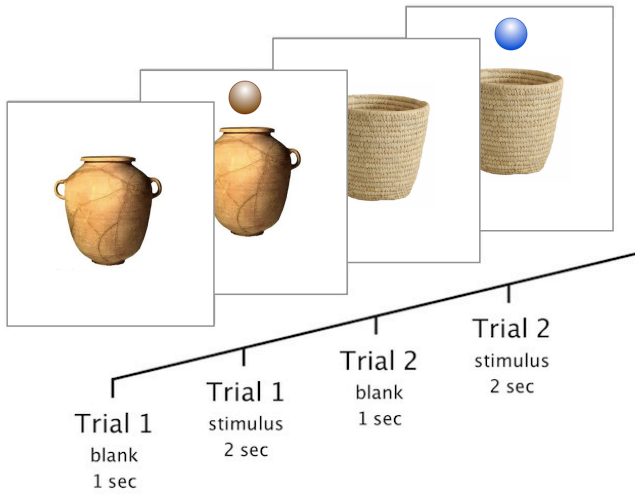


High load, observation phase



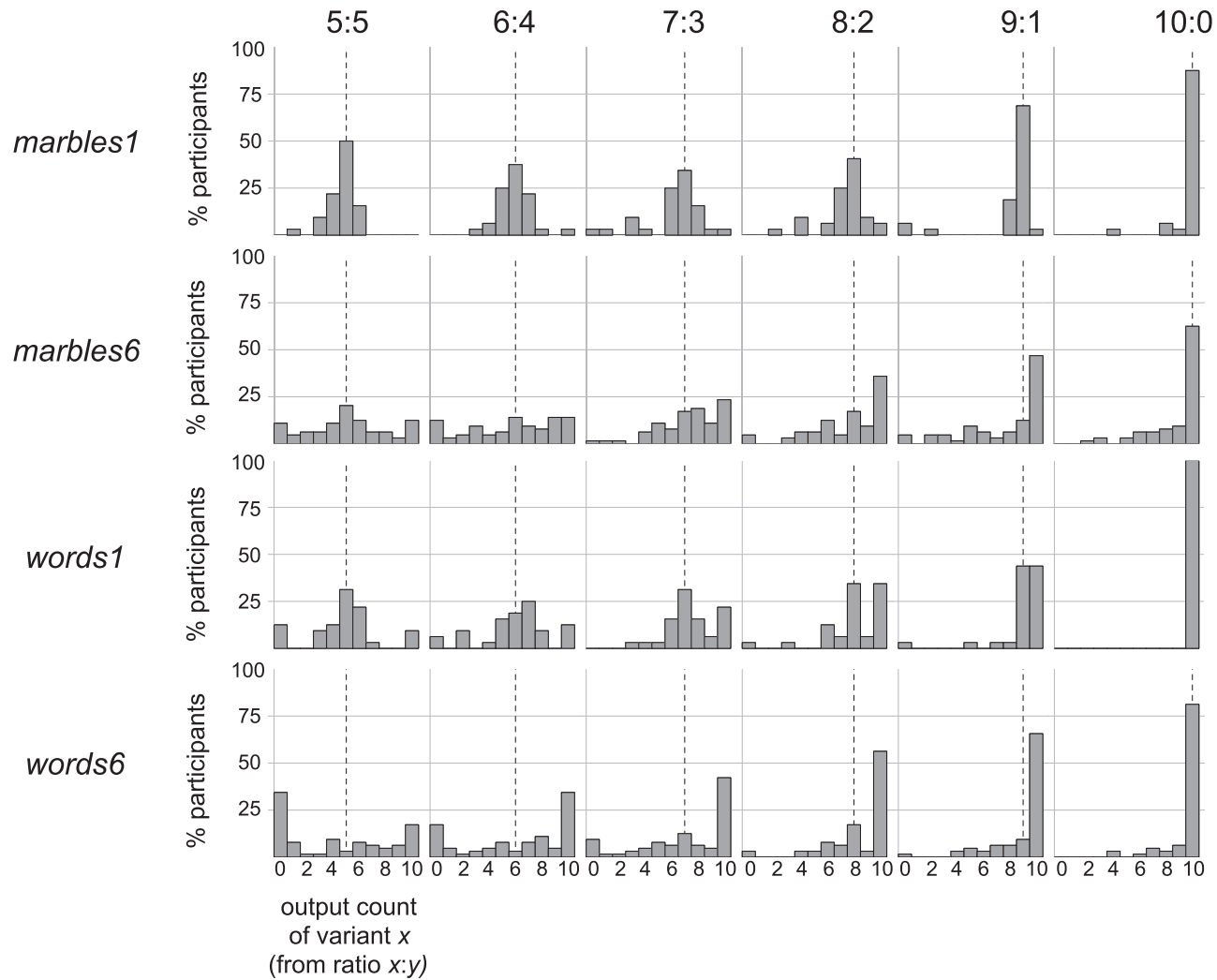
High load, production phase

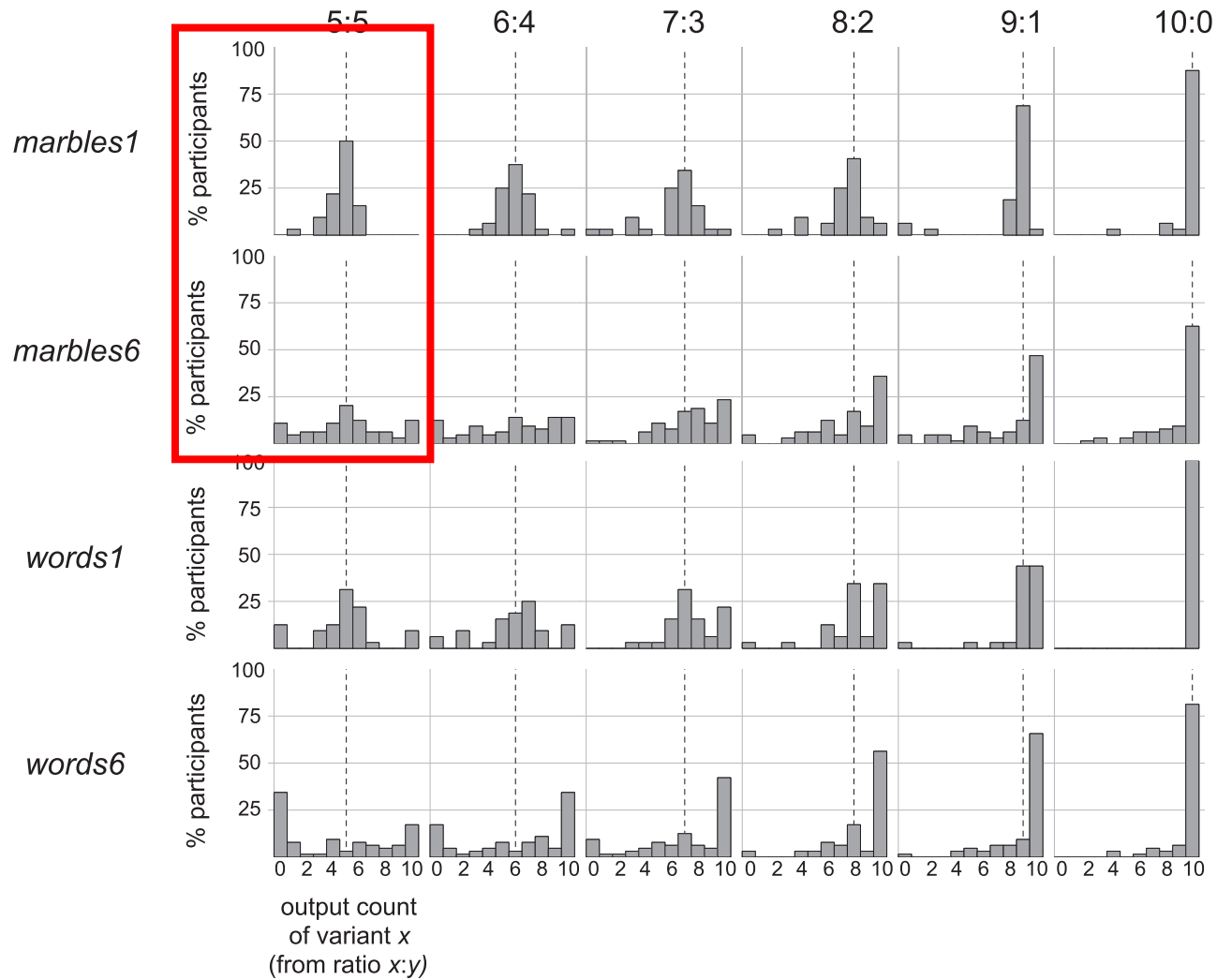


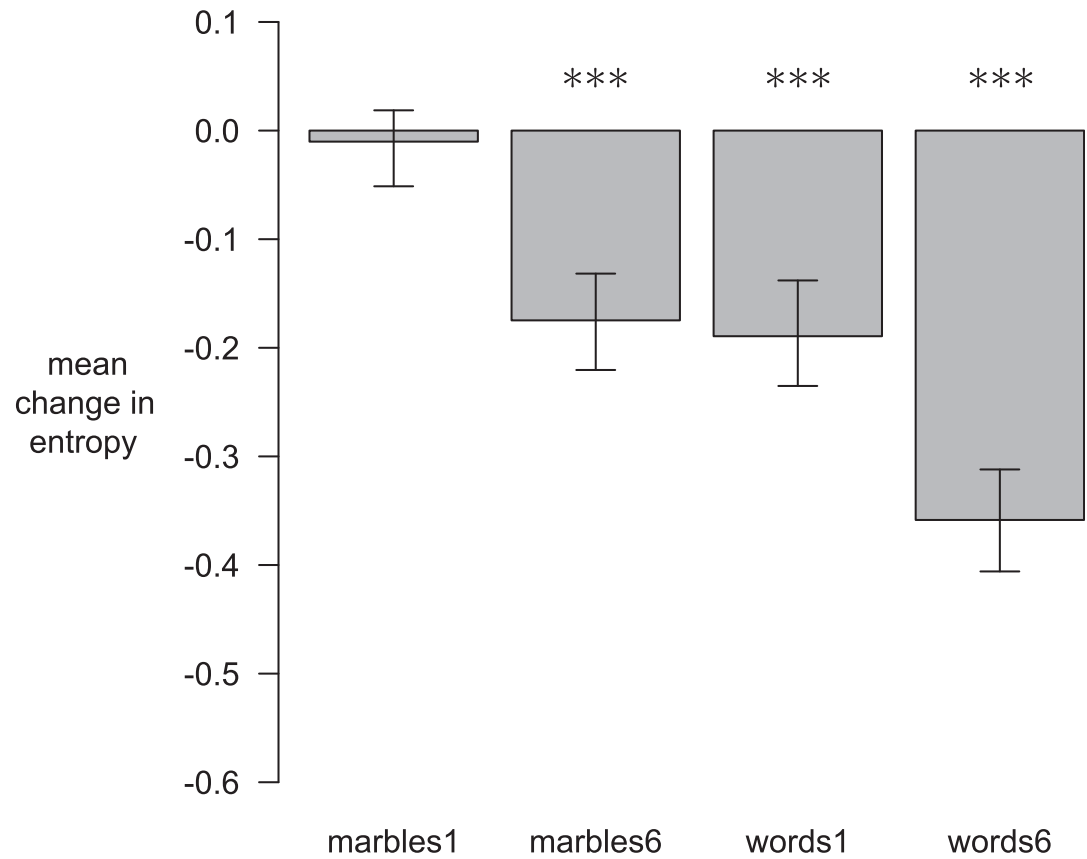


Sample size, study duration etc




- US-based MTurk workers
- N=512 after exclusions
- 4 minutes (1-item task) or 11.5 minutes (6-item task)
- **\$0.10** (1-item task) or **\$0.60** (6-item task) 😞

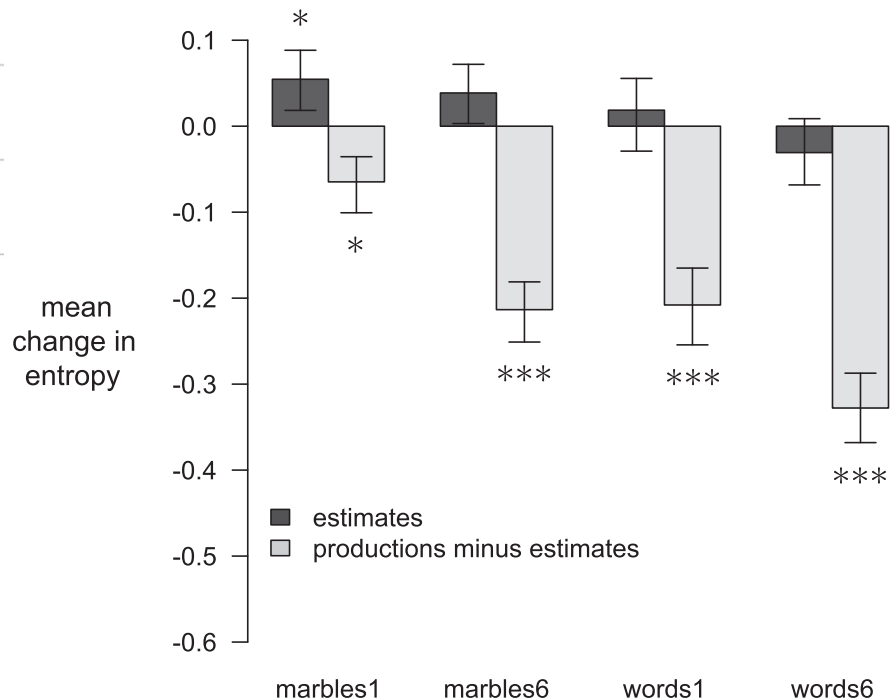






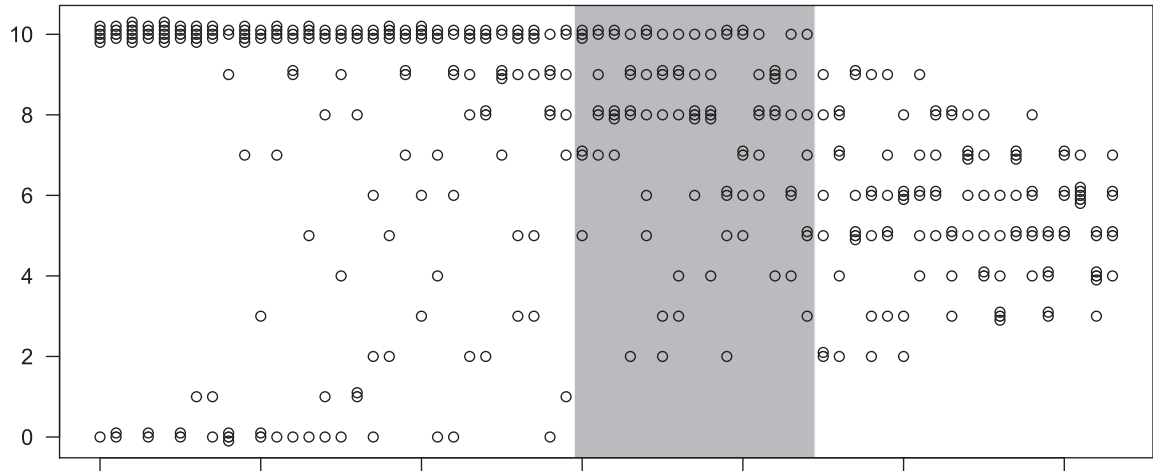
Regularization during encoding, or retrieval?

	tef	0	10	20	30	40	50	60	70	80	90	100
	gos	100	90	80	70	60	50	40	30	20	10	0
	fud	0	10	20	30	40	50	60	70	80	90	100
	pon	100	90	80	70	60	50	40	30	20	10	0
	seb	0	10	20	30	40	50	60	70	80	90	100
	nuk	100	90	80	70	60	50	40	30	20	10	0

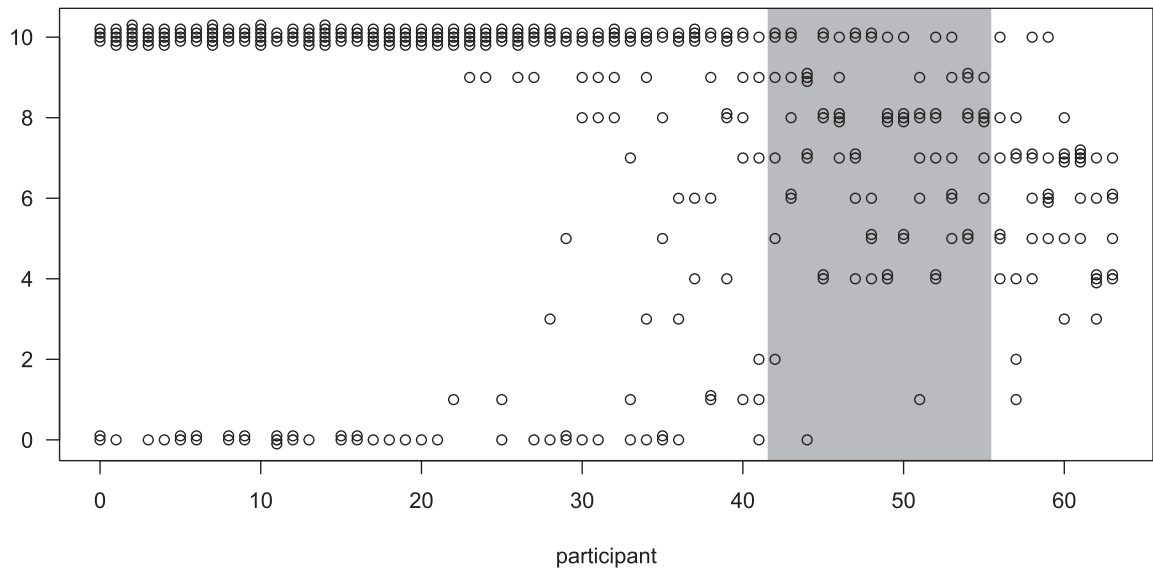


Individual differences

marbles6
output count
of variant x

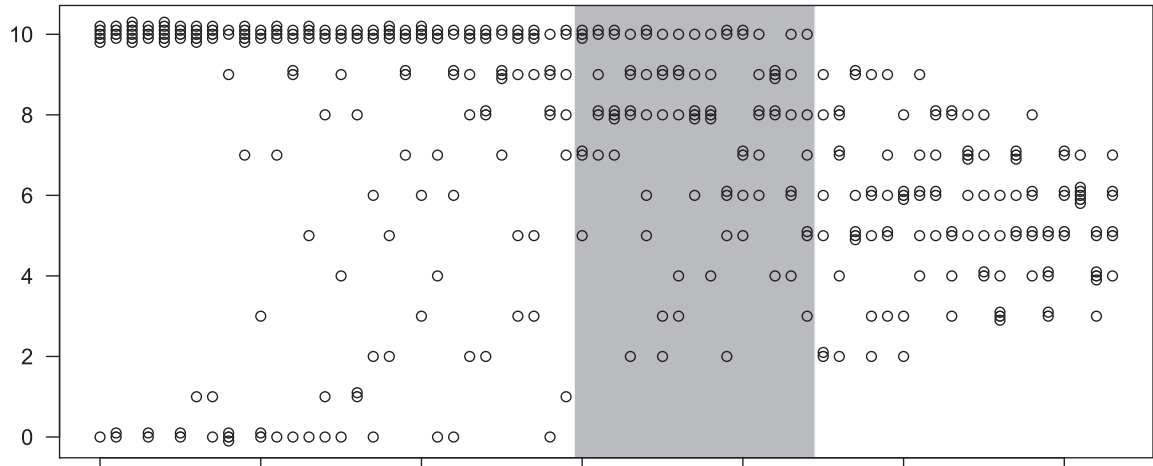


words6
output count
of variant x

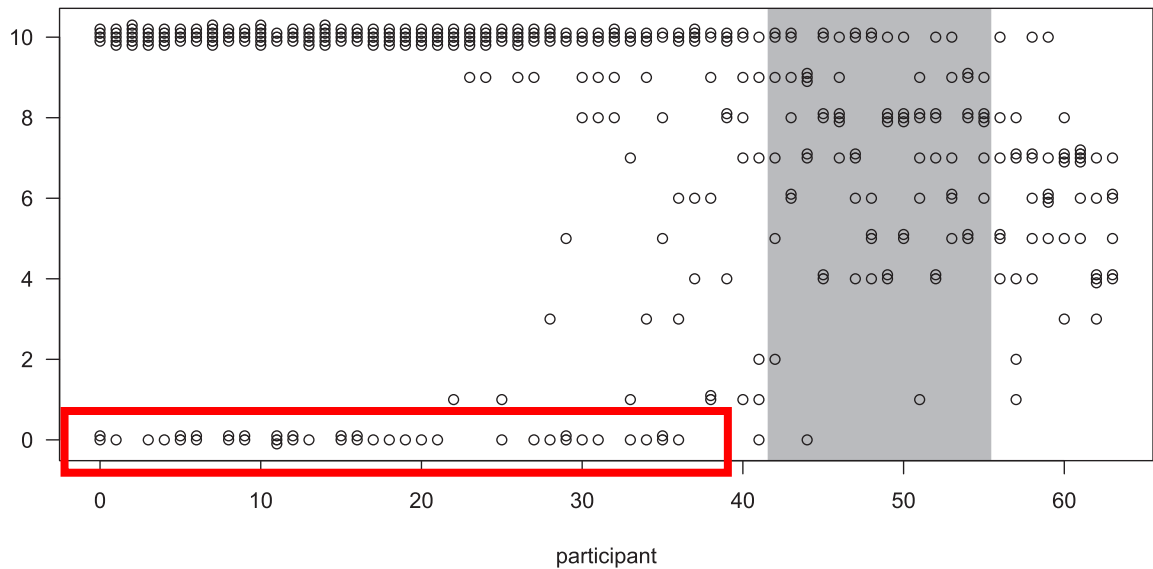


Minority regularizers

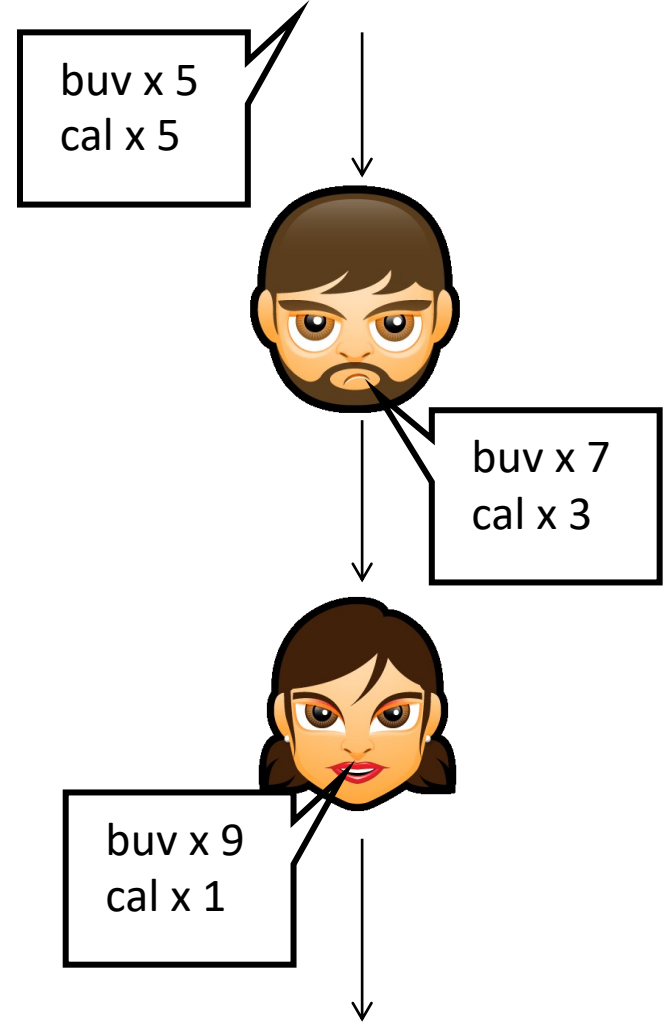
marbles6
output count
of variant x



words6
output count
of variant x



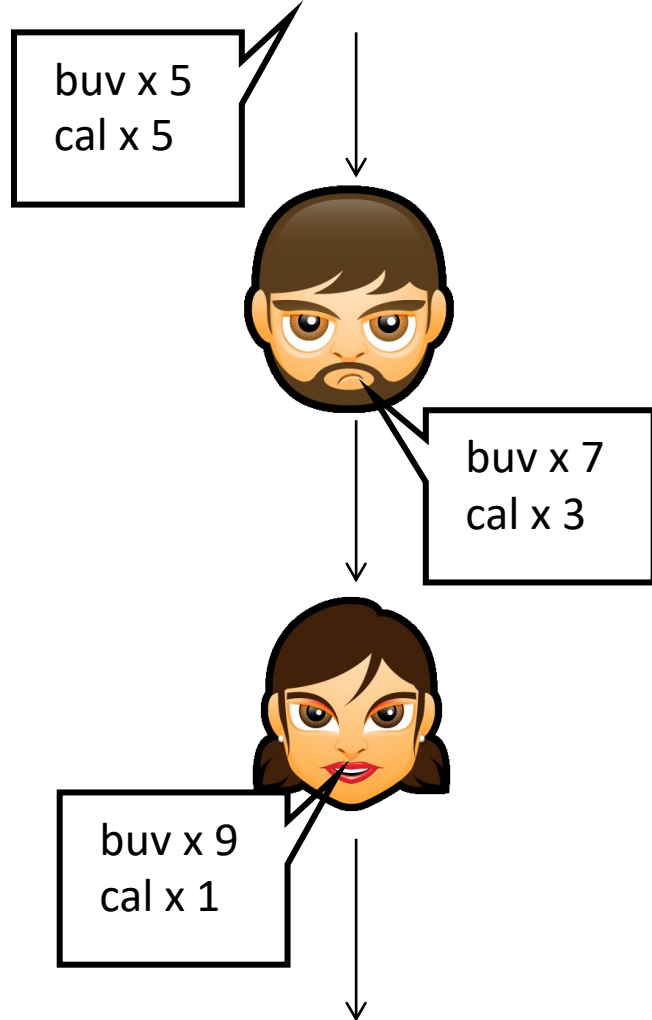
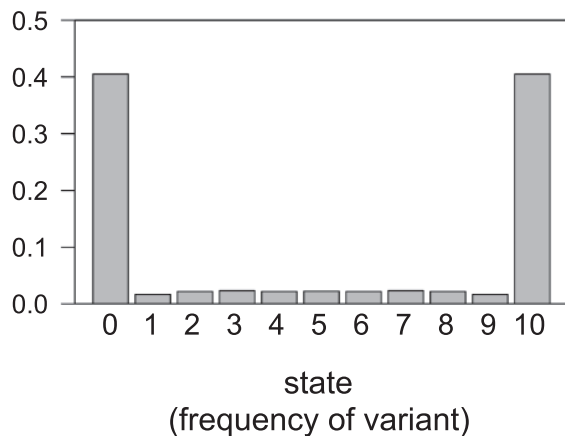
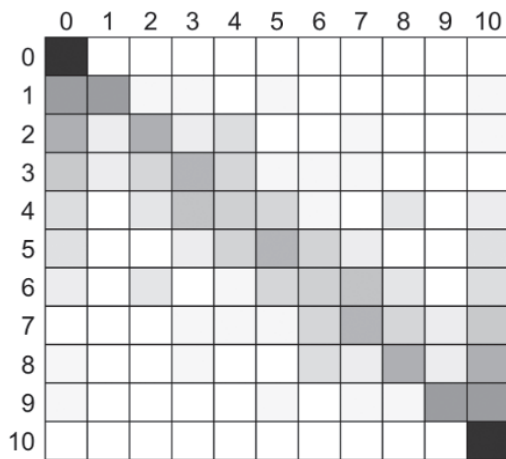
Simulating person-to-person transmission (iterated learning)



Simulating person-to-person transmission (iterated learning)

words1

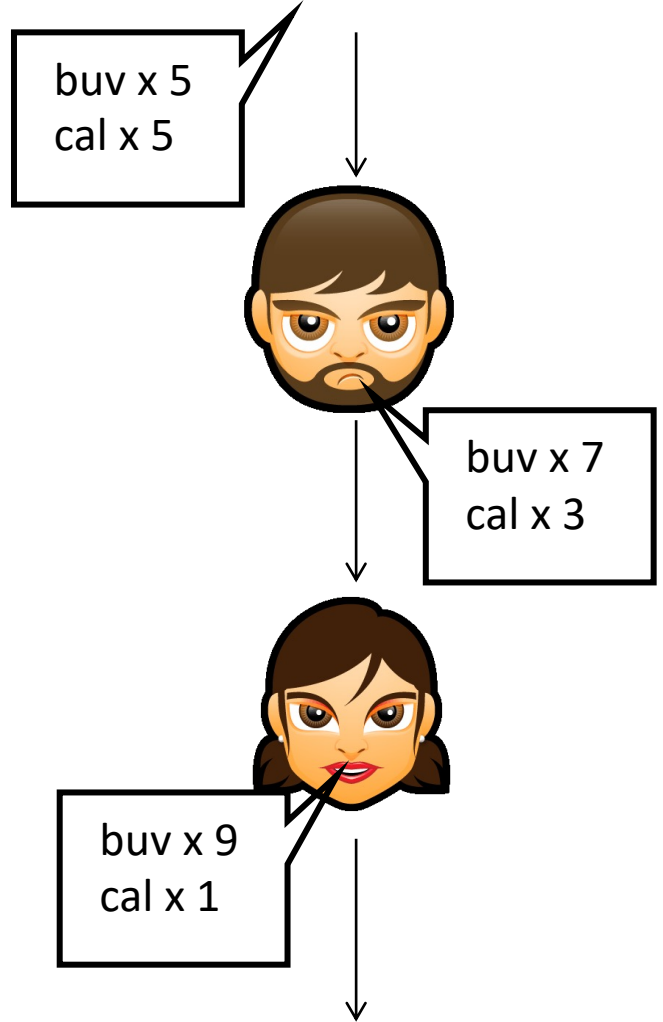
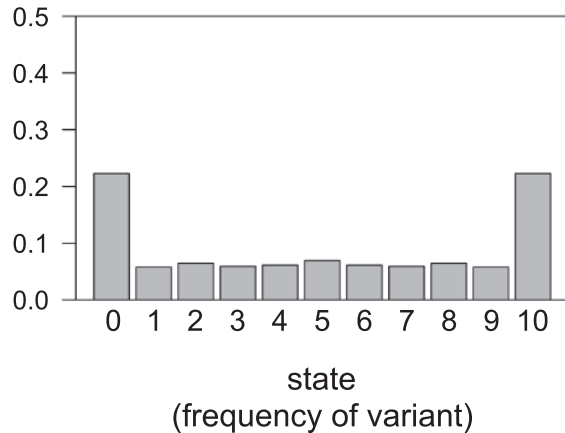
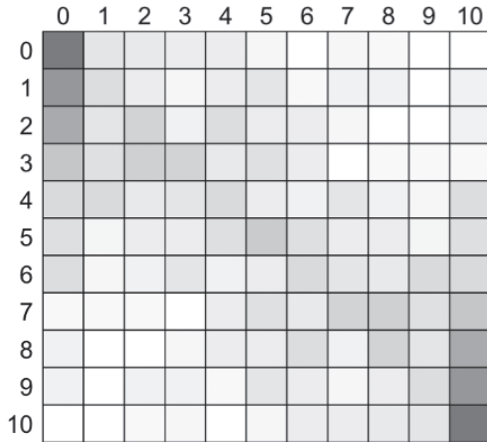
output frequency



Simulating person-to-person transmission (iterated learning)

marbles6

output frequency



Ferdinand et al.'s conclusions

Effects of domain and demand on regularization

- More regularization on linguistic than non-linguistic tasks (why?)
- More regularization when under greater cognitive load

Regularization effects mainly in recall (not encoding)

Simulation of iterated learning can reveal additional differences in regularization (cf. marbles₆ vs words₁)

Time for Q&A/discussion on this week's reading

Next up

Wednesday, 9am: lab

- A frequency learning experiment

Next week:

- Perceptual learning, audio stimuli