Online Experiments for Language Scientists

Lecture 5: Frequency learning

Kenny Smith

kenny.smith@ed.ac.uk

Reminder about Undergraduate Assessment 1

- Due on 9th November
- If you haven't already, read the assignment brief (https://kennysmithed.github.io/oels2023/assessment/UGAssignmentBrief2023.pdf) and FAQ (https://kennysmithed.github.io/oels2023/assessment/oels_assignment_faq.html)
- No questions after 10am on Monday 6th November!

Ferdinand, Kirby & Smith (2019)

Vanessa Ferdinand
(formerly Edinburgh, now Melbourne)

Simon Kirby
(Edinburgh)

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**?

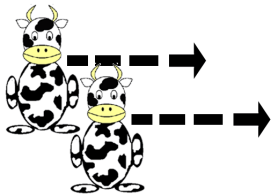
Variation in language

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

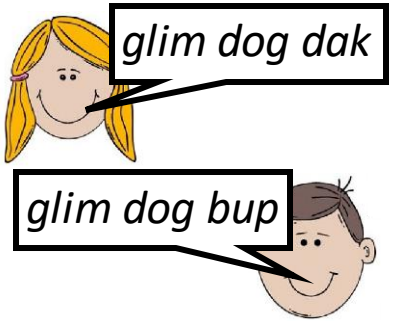
- Languages have lexicons and grammars
- Linguistic (phonological, lexical, syntactic, semantic) or sociolinguistic **conditioning** of alternation
  - English past tense allomorphy: hunt**/ɪd/** vs fish**/t/**
  - Noun classes: ***la*** *chaise,* ***le*** *sofa,* ***la*** *fille,* ***le*** *garçon*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

Why is language like this?

Variation-learning experiments

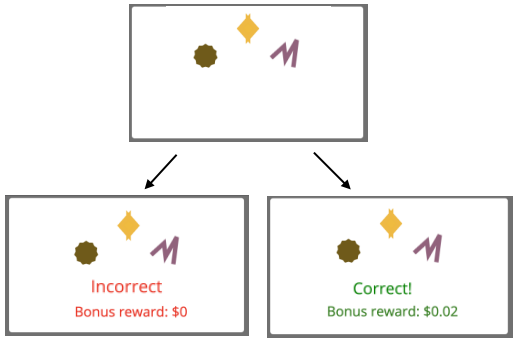


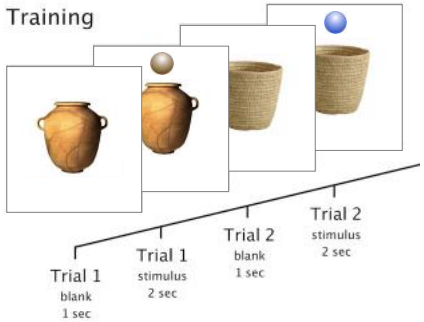
glim cow fip
glim cow tay



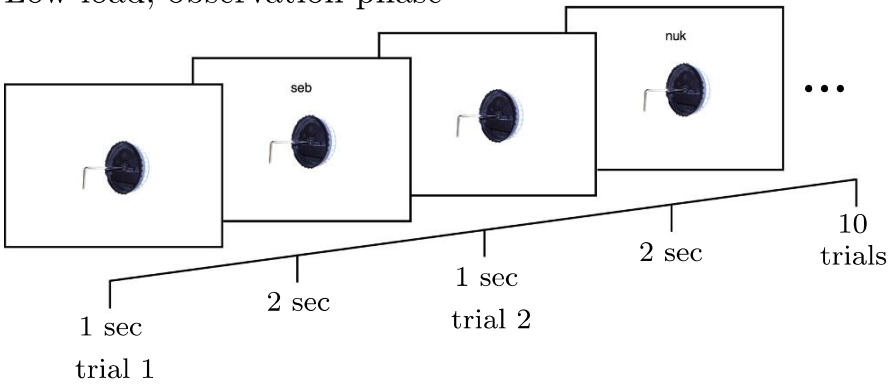
ooshra buzzo trunko
ooshra trunko tid buzzo

buv
kal

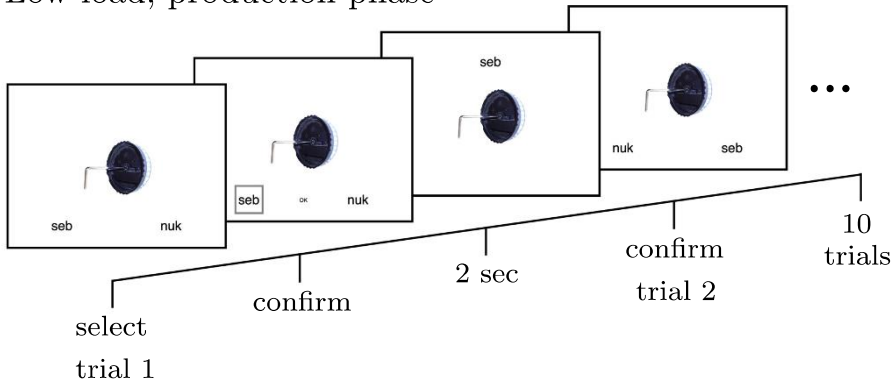




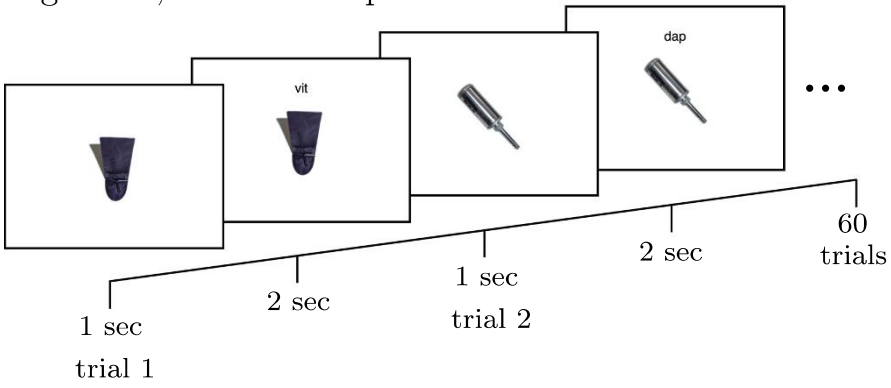
Low load, observation phase



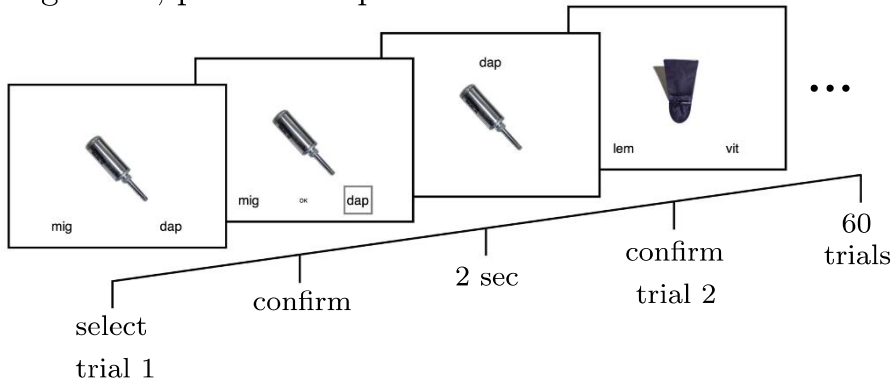
Low load, production phase

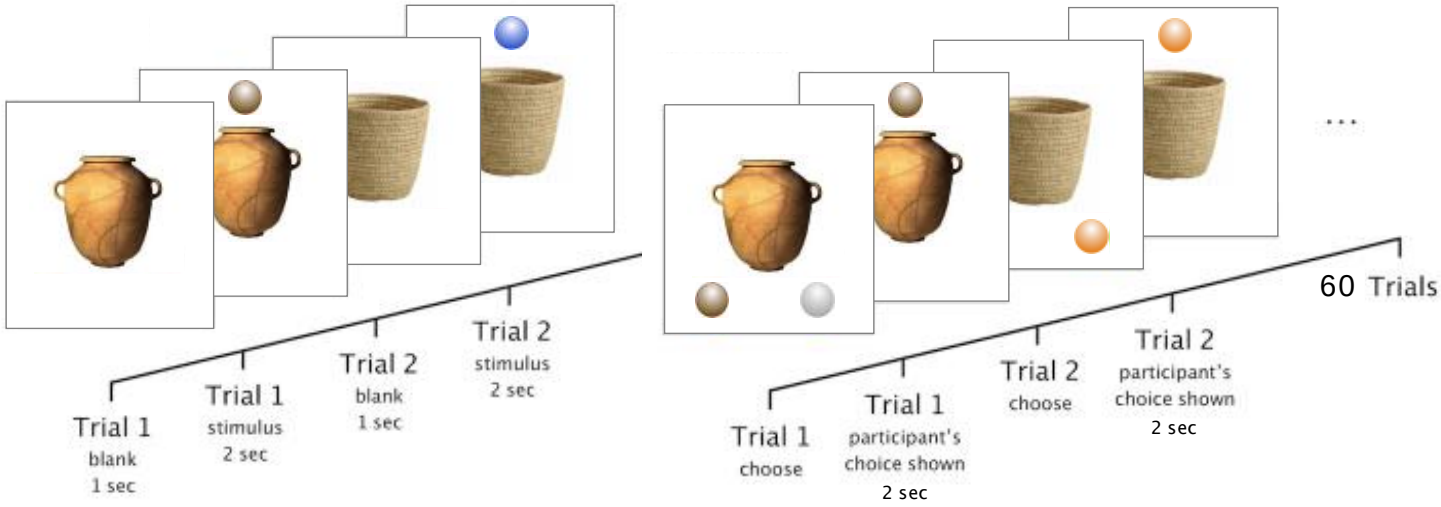


High load, observation phase



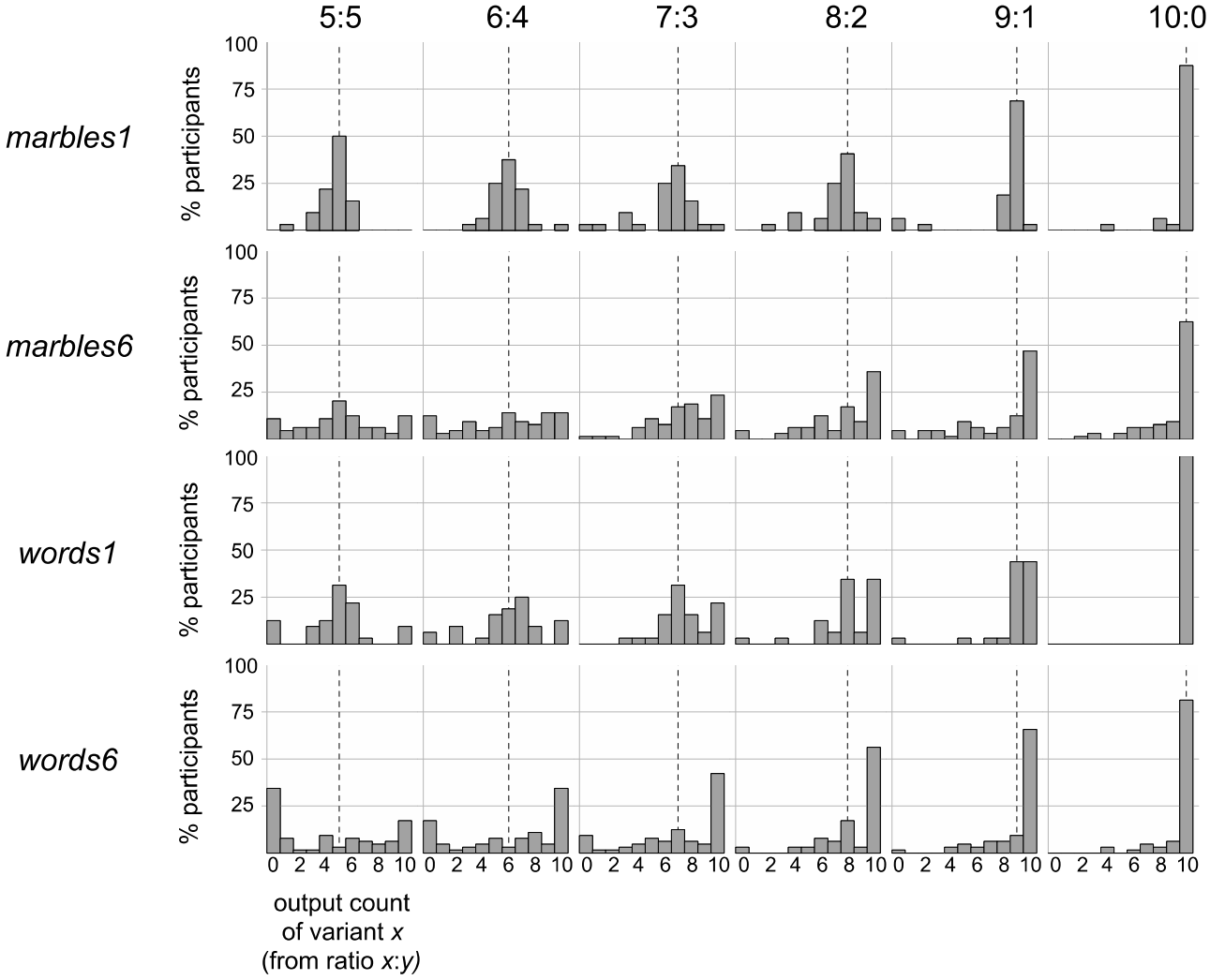
High load, production phase

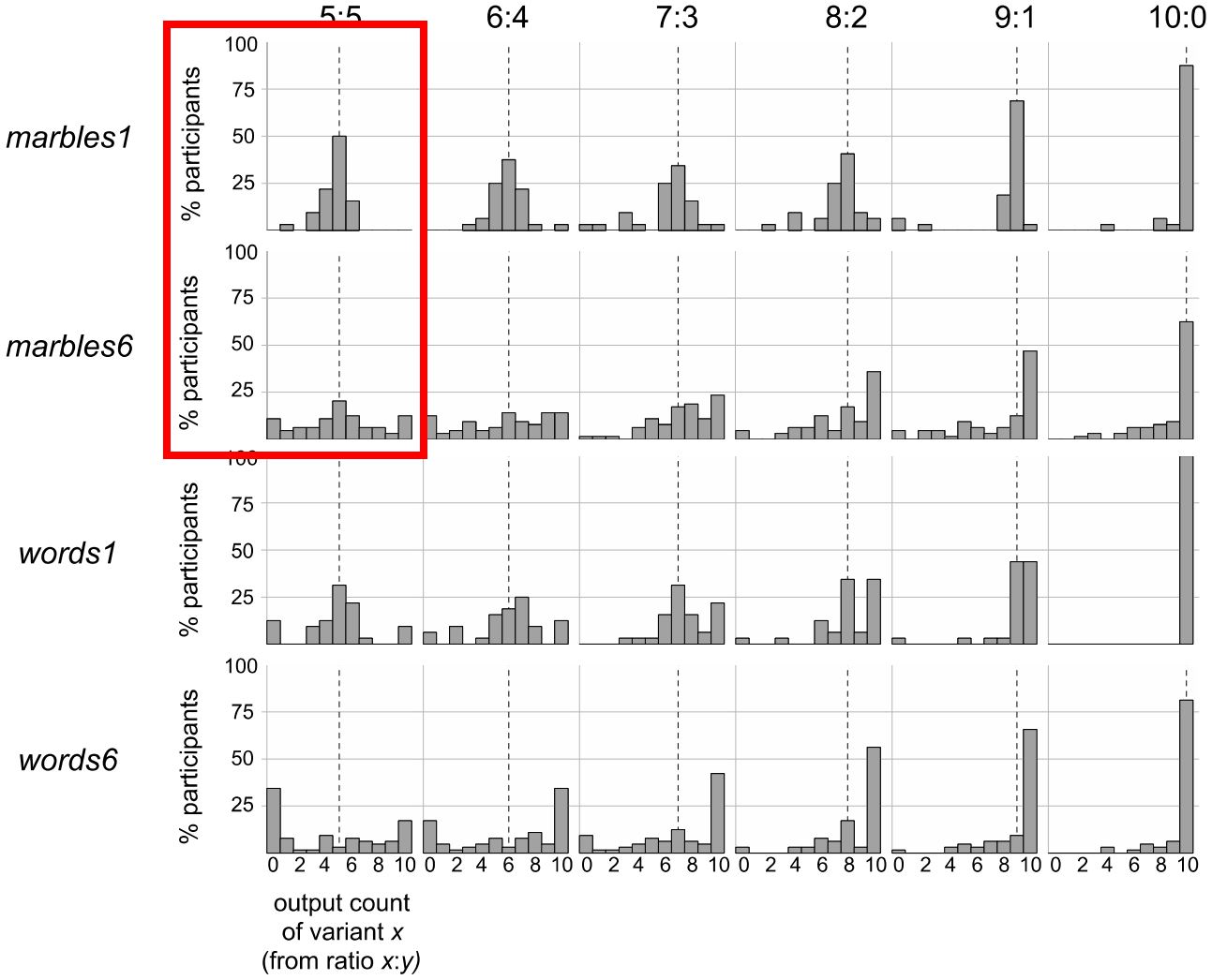


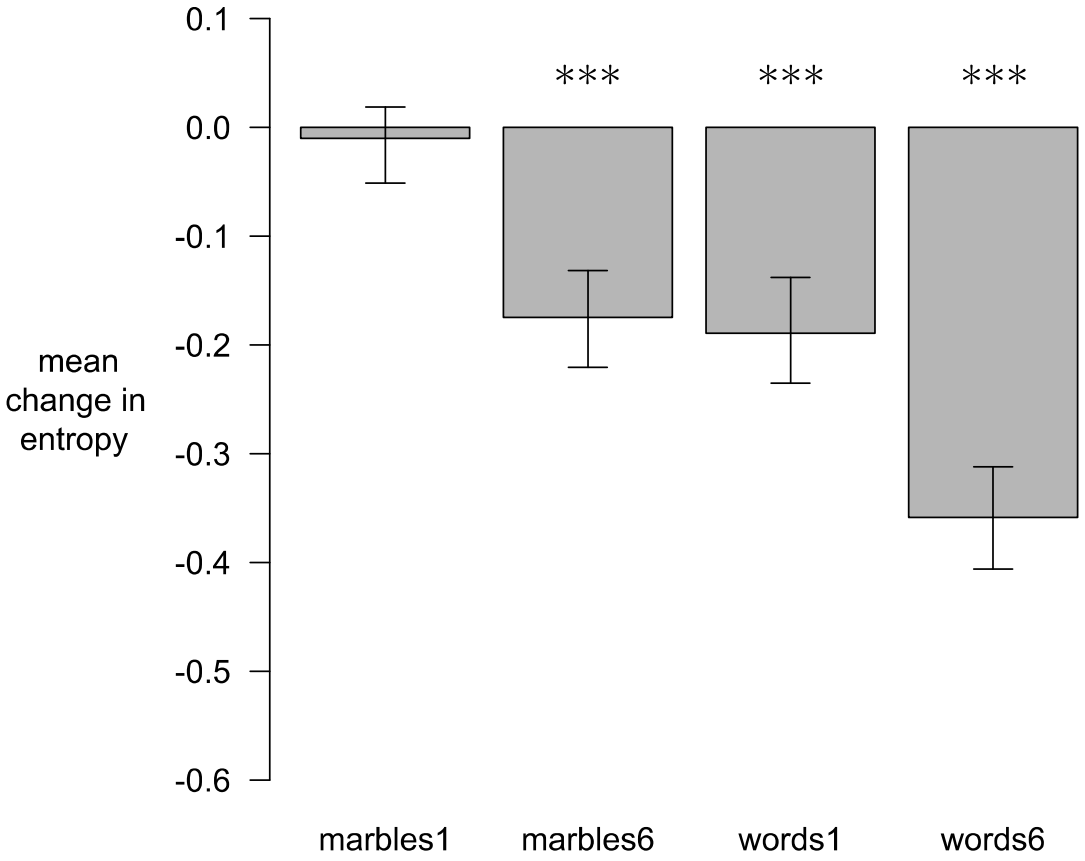

Trial 1
blank
1 sec
Trial 1
stimulus
2 sec
Trial 2
blank
1 sec
Trial 2
stimulus
2 sec
Trial 1
choose
Trial 1
participant's
choice shown
2 sec
Trial 2
choose
Trial 2
participant's
choice shown
2 sec
...
60 Trials

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) 😢


5:5
6:4
7:3
8:2
9:1
10:0
marbles1
marbles6
words1
words6
% participants
100
75
50
25
0 2 4 6 8 10
output count
of variant x
(from ratio x:y)


5:5
6:4
7:3
8:2
9:1
10:0
marbles1
marbles6
words1
words6
% participants
100
75
50
25
0 2 4 6 8 10
output count
of variant x
(from ratio x:y)

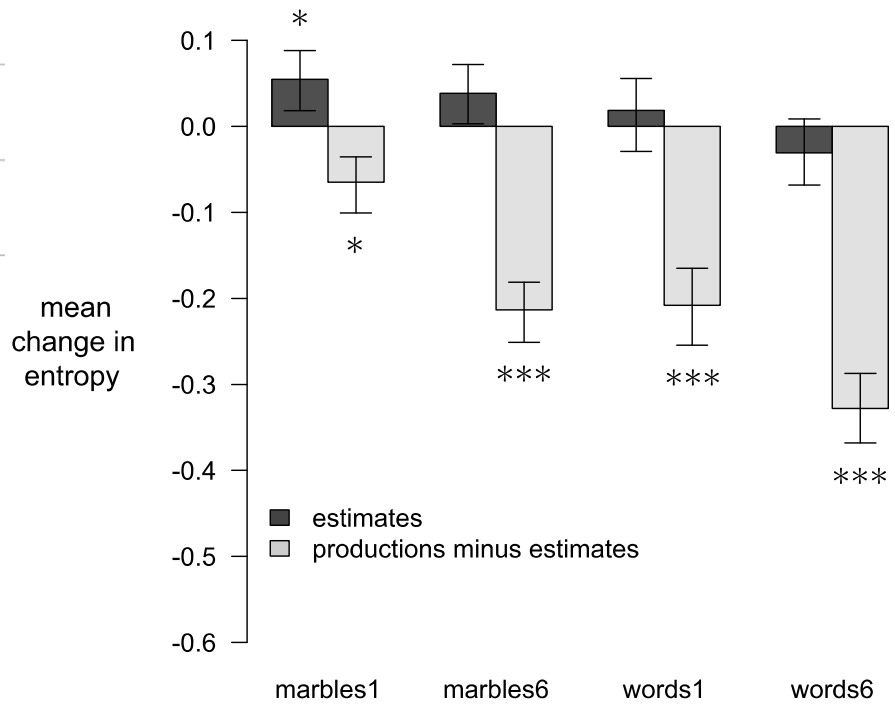

0.1
0.0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
mean
change in
entropy
***
***
***
marbles1
marbles6
words1
words6

Regularization during encoding, or retrieval?

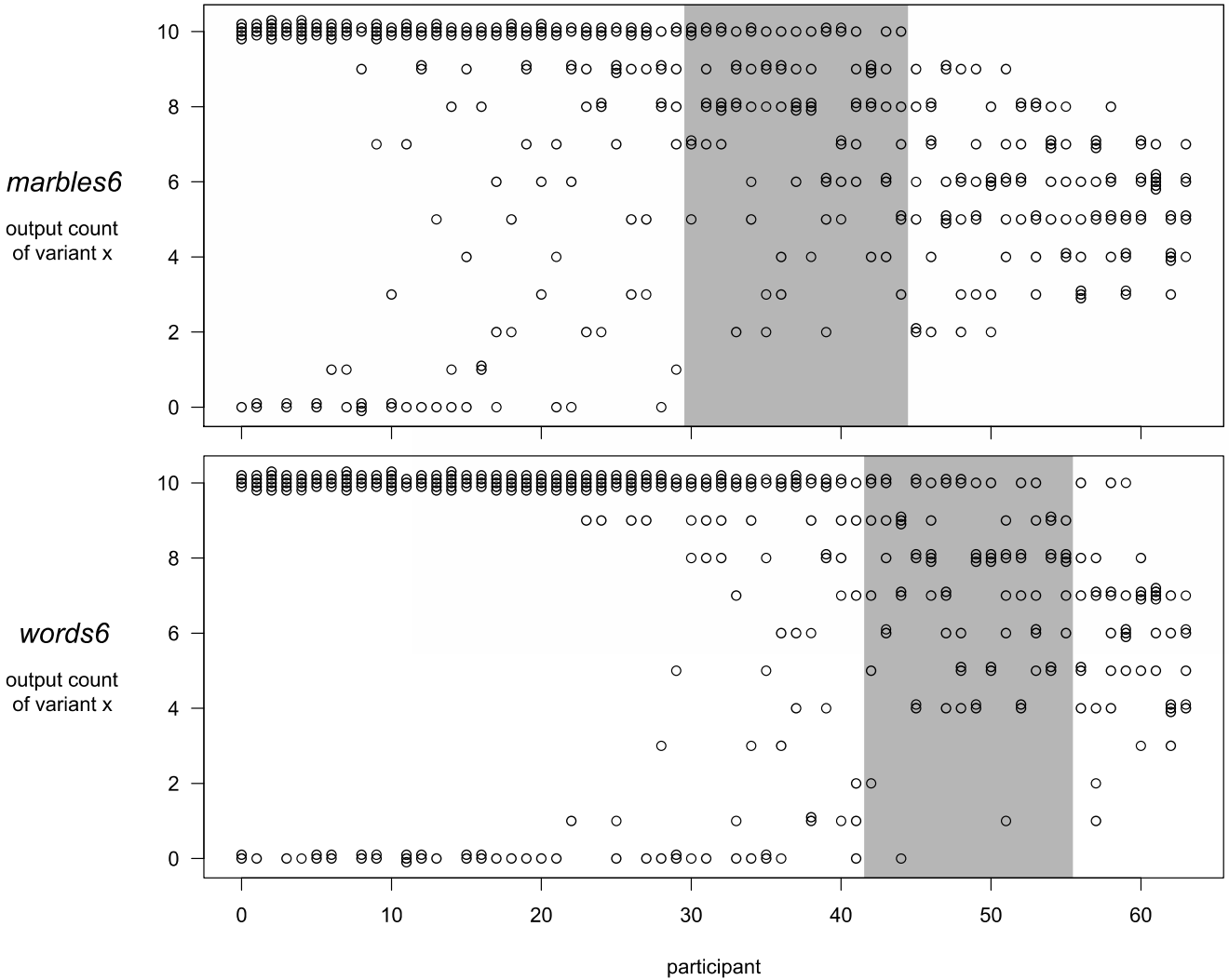
tef	0	10	20	30	40	**50**	60	70	80	90	100
gos	100	90	80	70	60	**50**	40	30	20	10	0

fud	0	10	20	30	40	50	60	70	80	**90**	100
pon	100	90	80	70	60	50	40	30	20	**10**	0

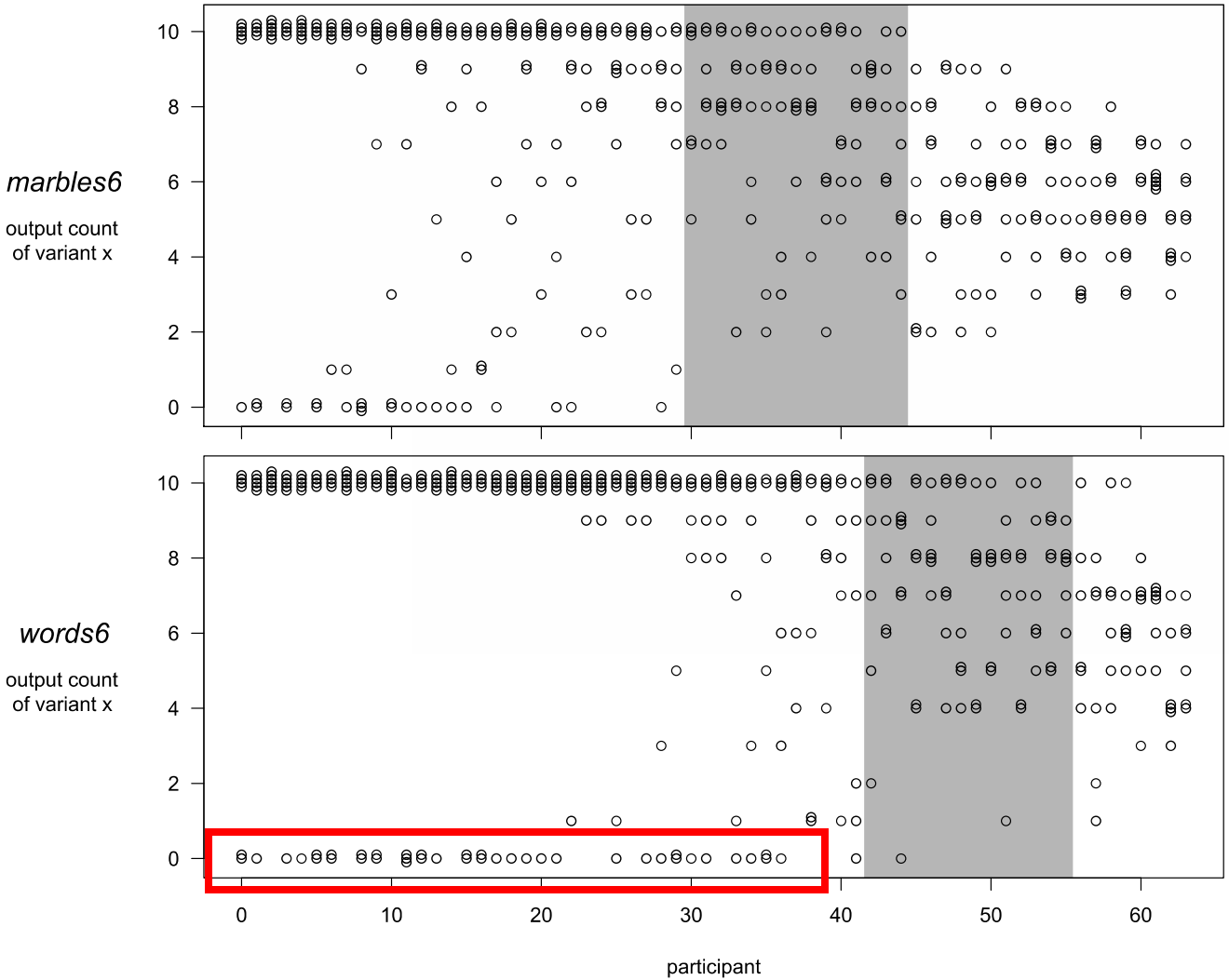
seb	0	10	**20**	30	40	50	60	70	80	90	100
nuk	100	90	**80**	70	60	50	40	30	20	10	0



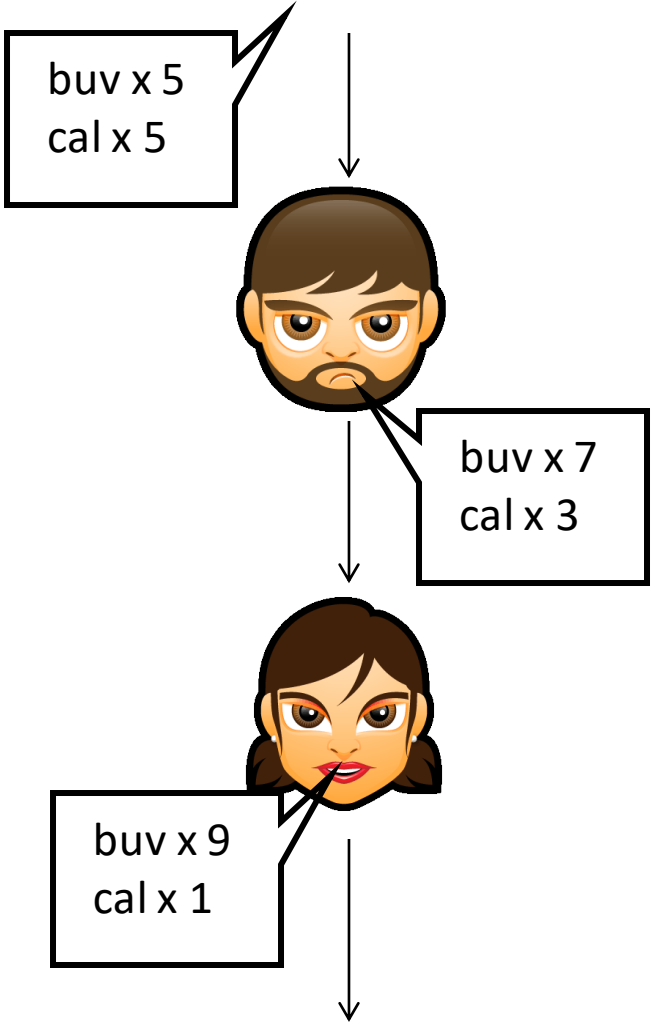
Individual differences



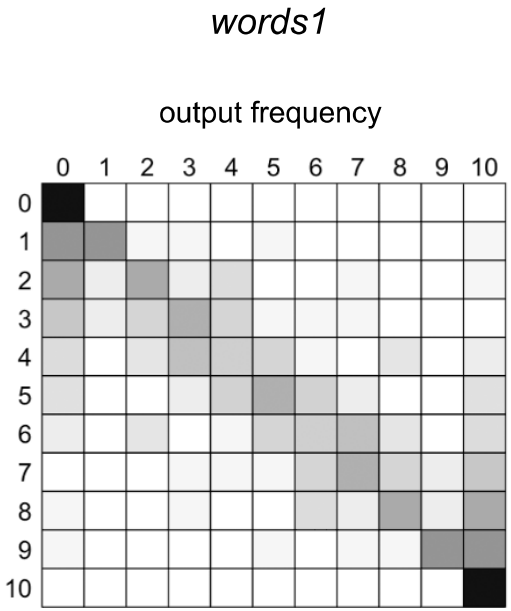
Minority regularizers

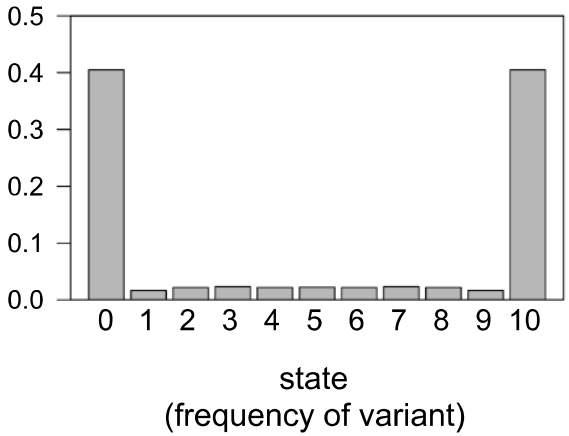


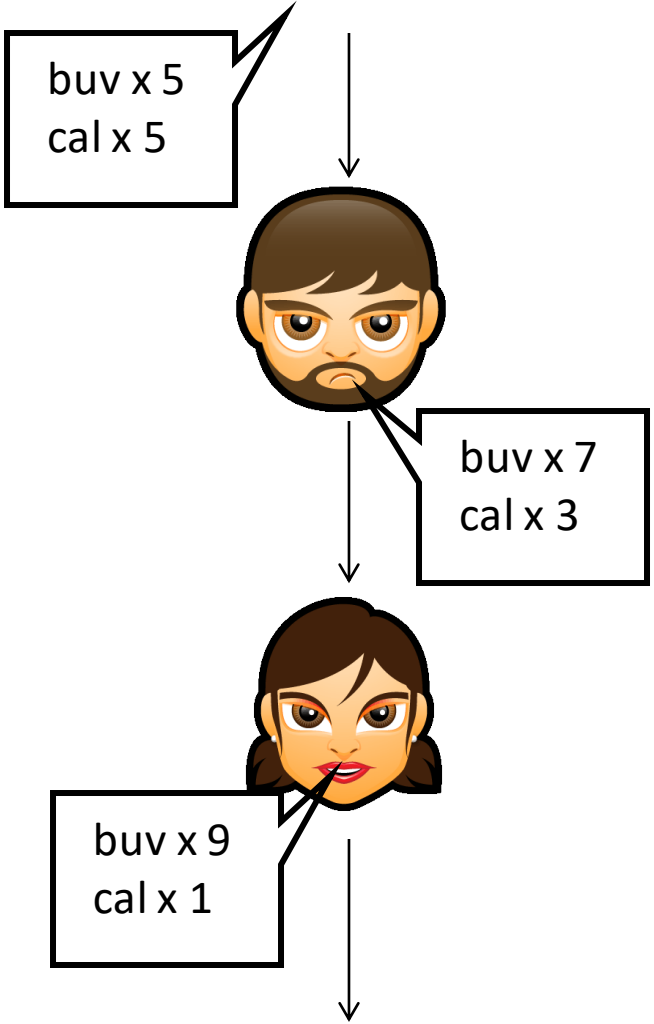
Simulating person-to-person transmission (iterated learning)



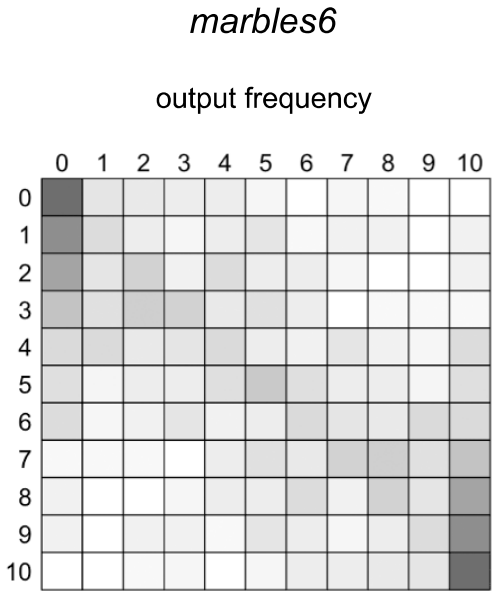
Simulating person-to-person transmission (iterated learning)

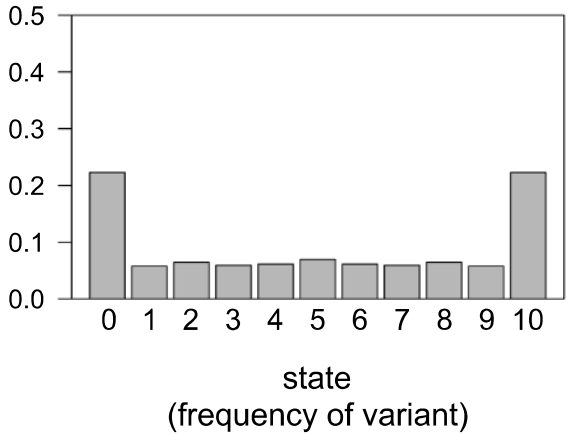


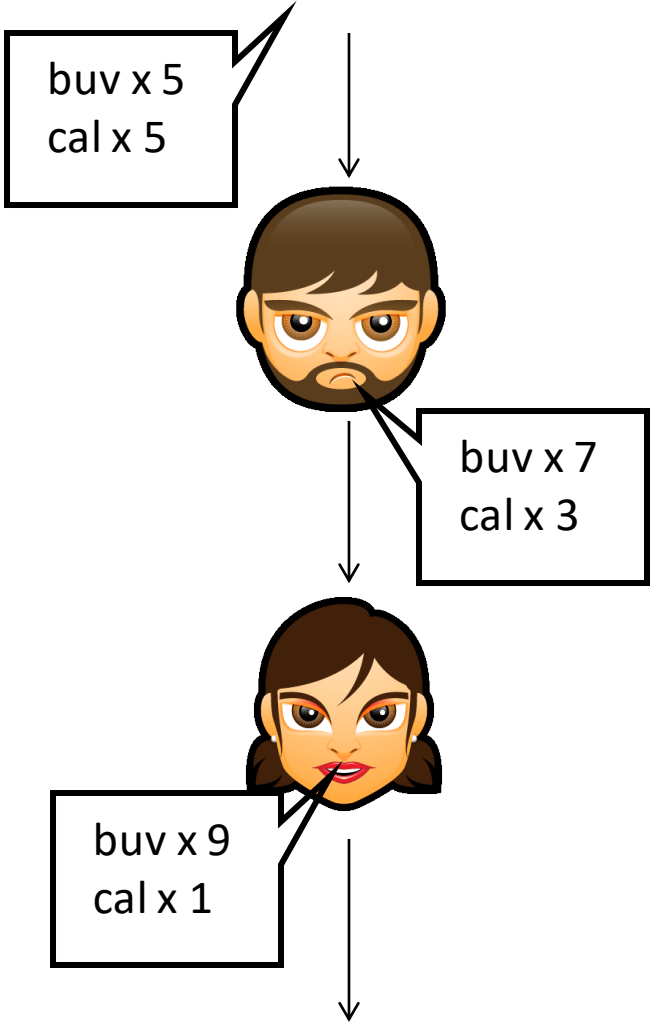




Simulating person-to-person transmission (iterated learning)







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

Lab

- A frequency learning experiment

Next week:

- Perceptual learning, audio stimuli
- The end of the “basics”