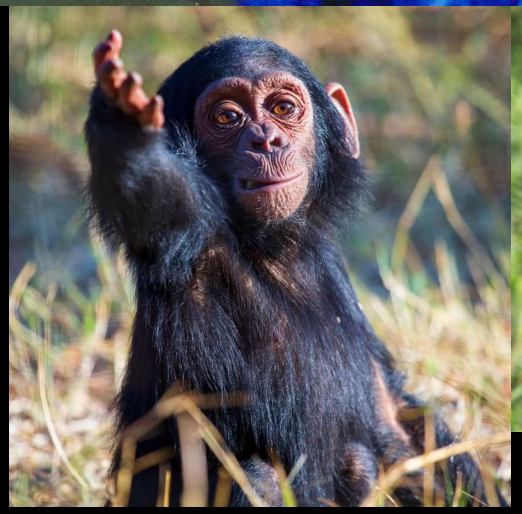
The evolution of linguistic regularities and exceptions

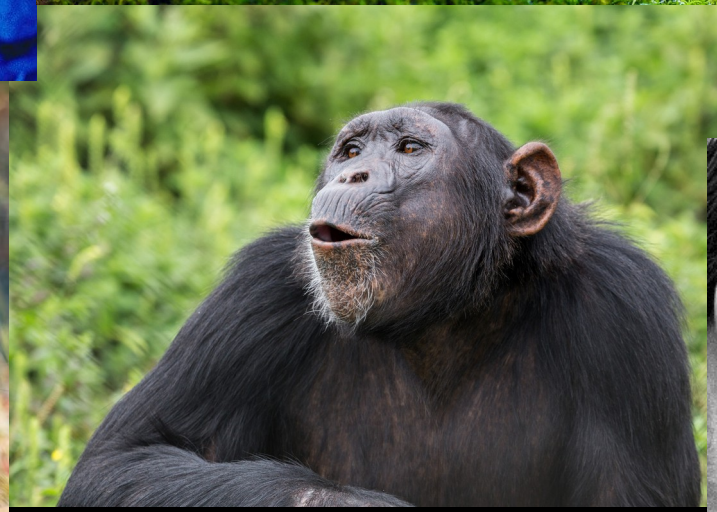
Kenny Smith

kenny.smith@ed.ac.uk

erc

CENTRE for
LANGUAGE EVOLUTION
www.lel.ed.ac.uk/cle





Communication is widespread,
but language is unique

Language’s open-ended communicative power comes from its **regular compositional structure**


John loves Mary


Yesterday I saw a
pink cat!

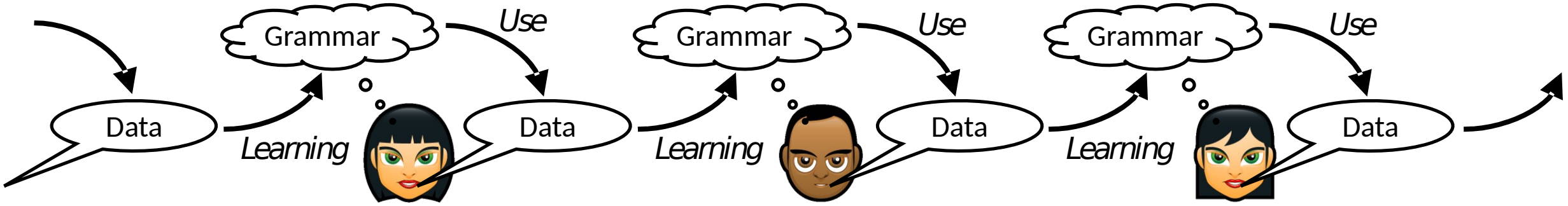

A Russian woman was sentenced to 21 years for trying to kill her doppelganger with a poisoned cheesecake!!!
?
?

Language is transmitted via repeated **learning** and **use**

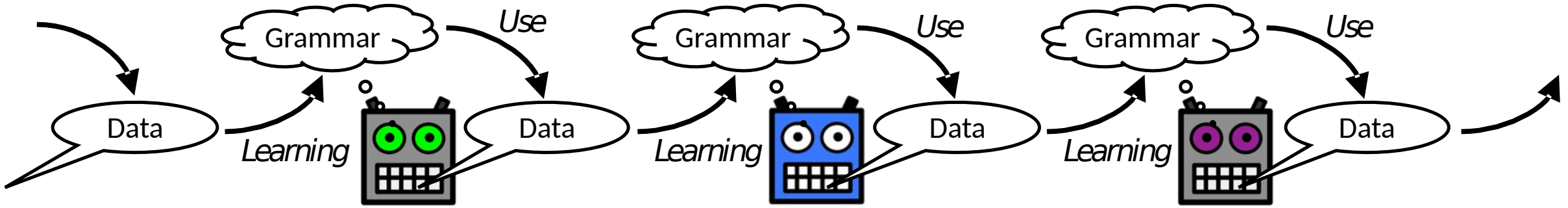
Language is shaped by these processes

These processes create create the regular compositional structure that makes language so powerful

And we can understand **exceptions** to those regularities (within and across languages) in terms of those same processes

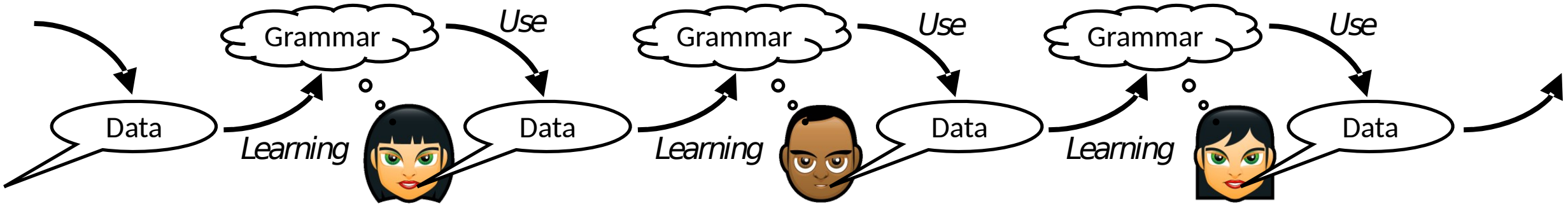


From Smith, 2022, *Current Directions in Psychological Science*
(based on e.g. Hurford, 1990, in *Logical Issues in Language Acquisition*)



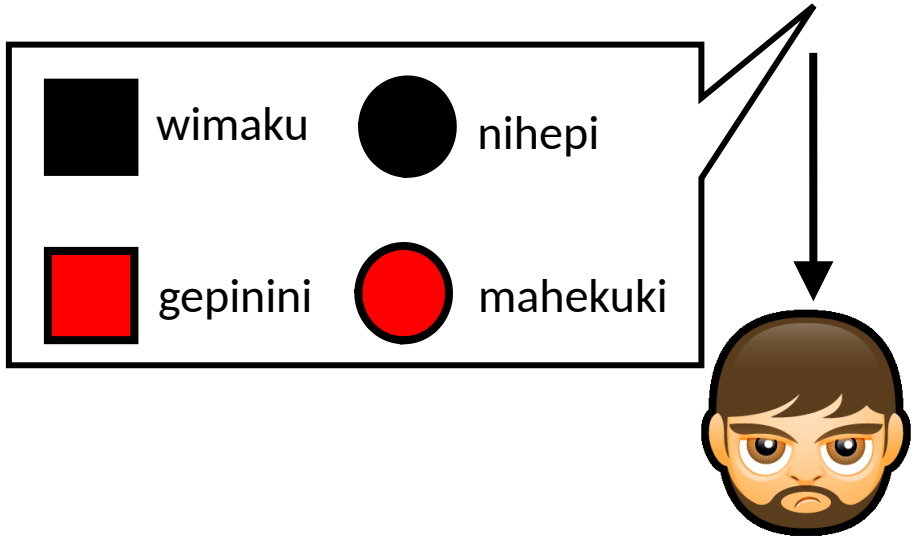
Computational experiments

See e.g. Kirby, Griffiths, & Smith (2014). *Current Opinion in Neurobiology* for review

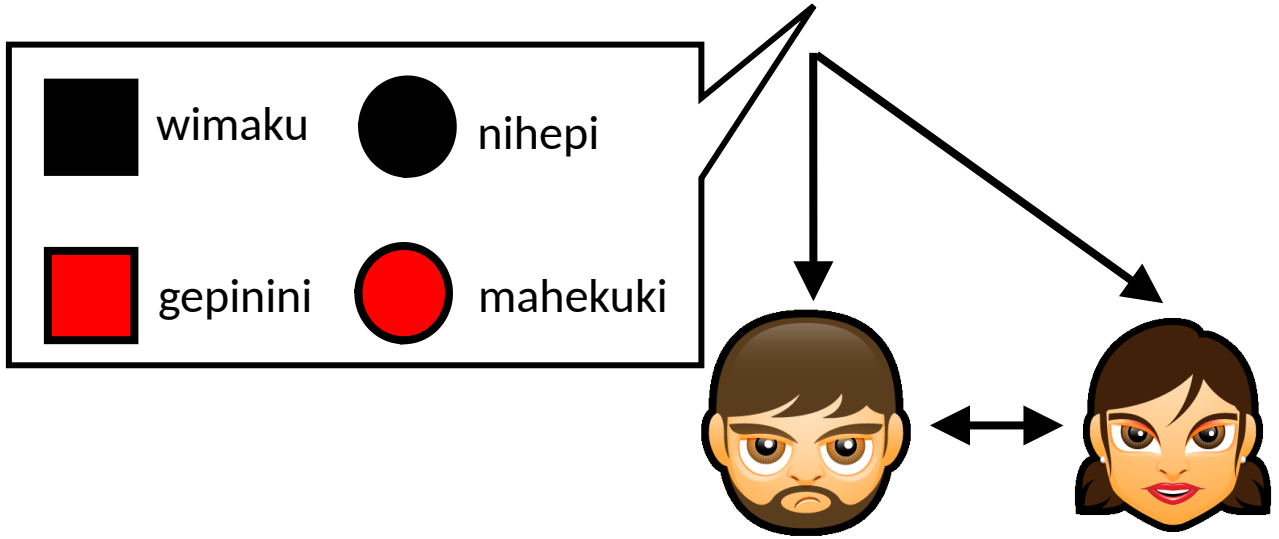


Lab experiments

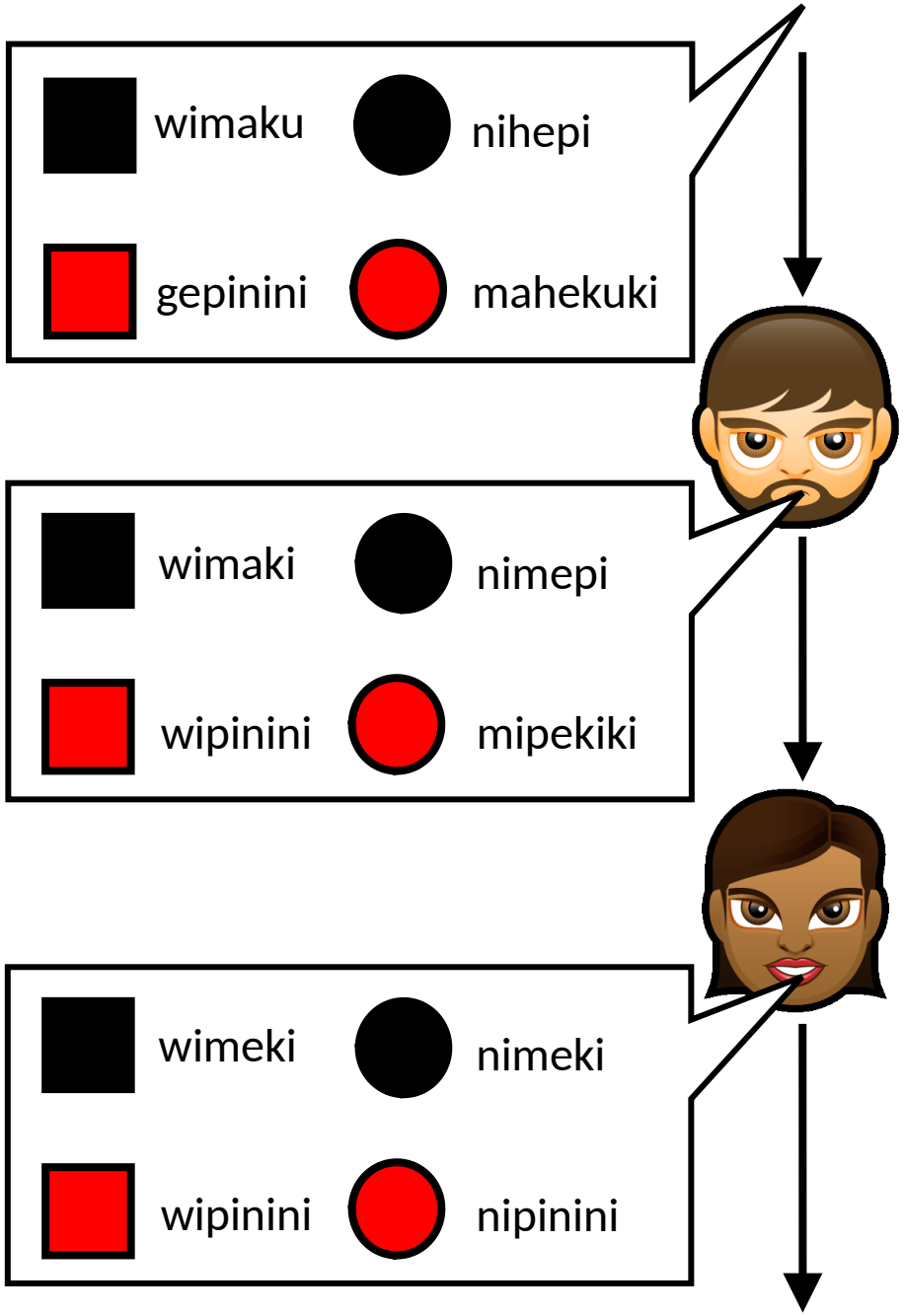
See e.g. Smith (2022). *Current Directions in Psychological Science* for review



Artificial Language Learning



Communicative interaction



Iterated learning

Talk outline

Part 1: learning and use create regular compositional structure

Part 2: learning also explains where irregularity appears **within** languages (high-frequency things are more likely to be irregular)

Part 3: learning also explains where irregularity appears **across** languages (languages with fewer non-native speakers are more complex and irregular)

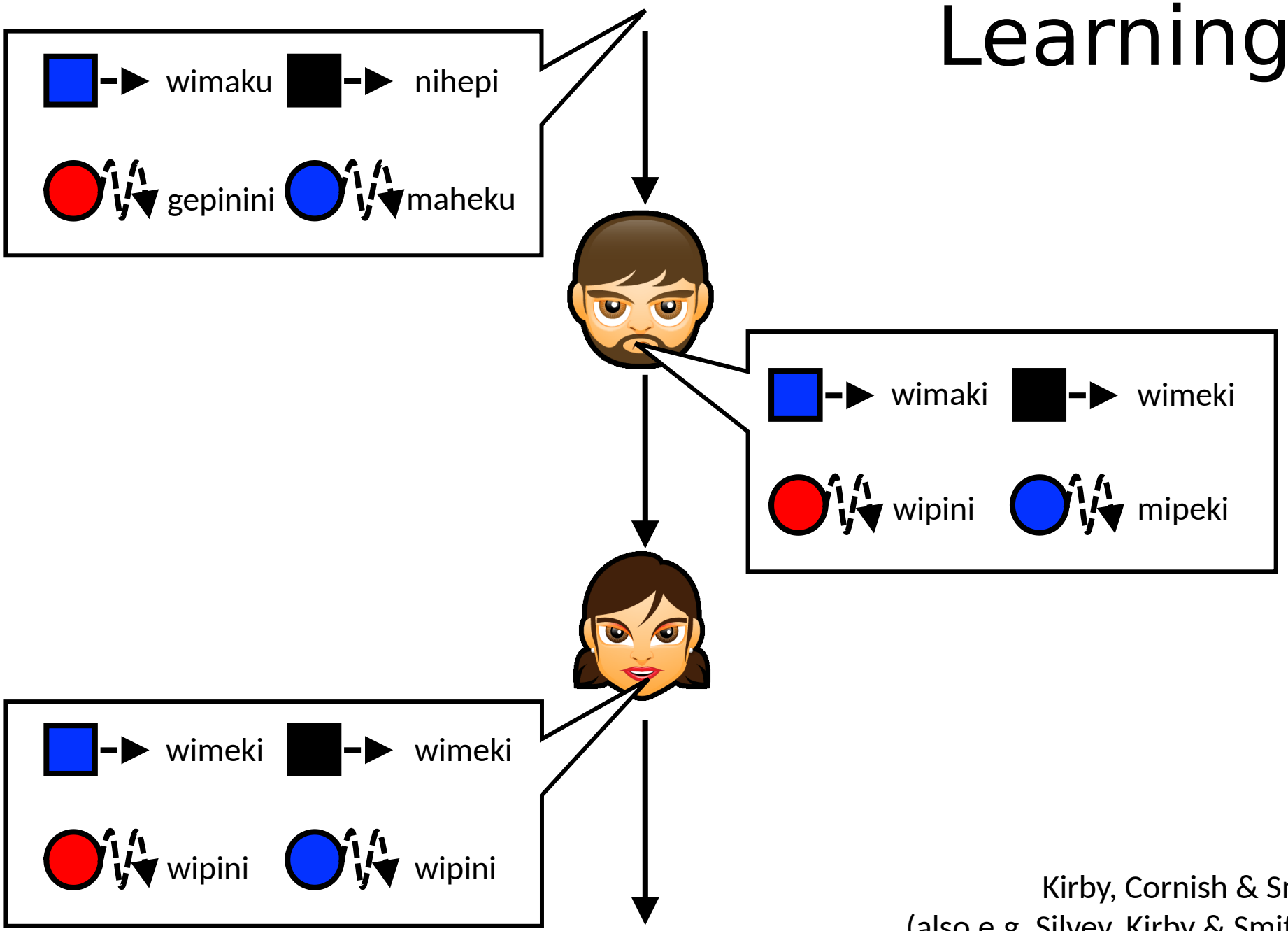
Part 1: learning and use create regular compositional structure

Simon Kirby
(Edinburgh)

Monica Tamariz
(Heriot Watt)

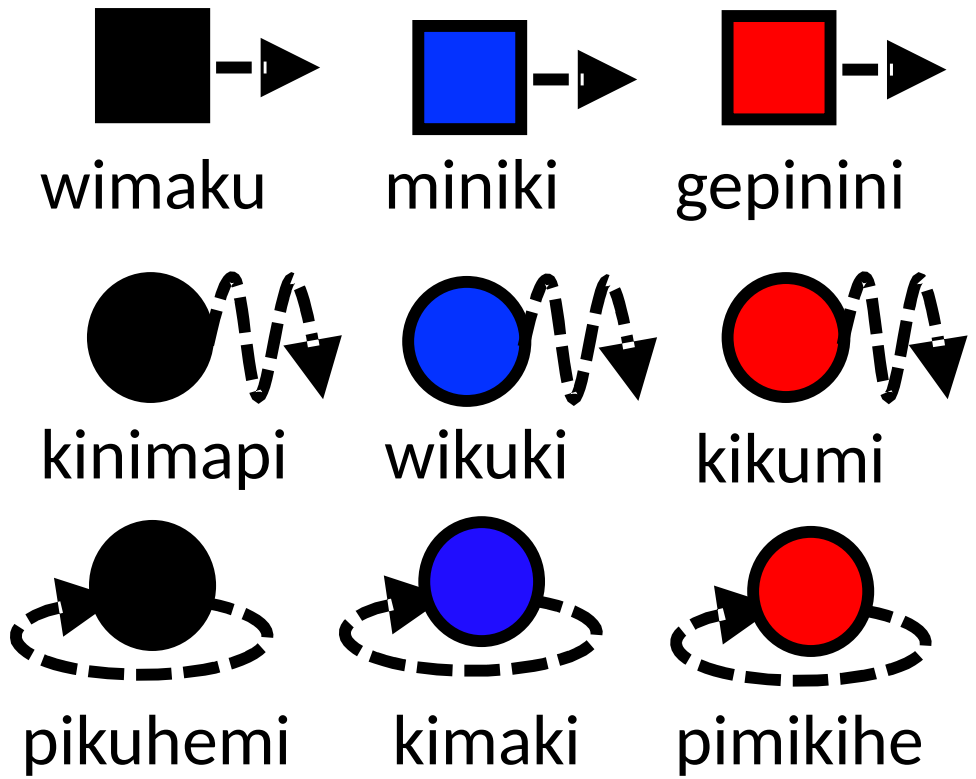
Hannah Cornish
(Edinburgh)

Learning only

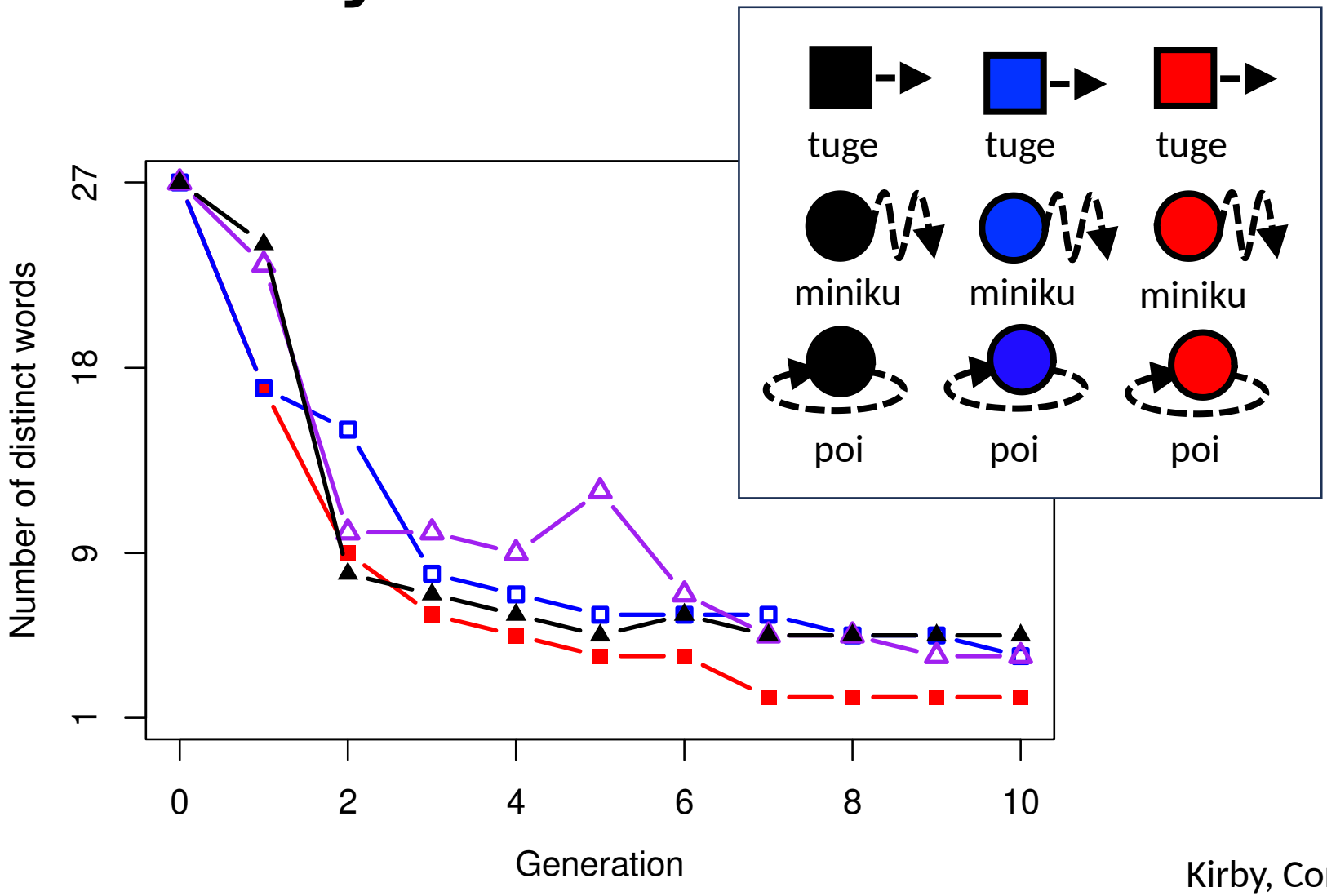


Kirby, Cornish & Smith, 2008, *PNAS*
(also e.g. Silvey, Kirby & Smith, 2015, *Cog Sci)*

Starting point: non-compositional ‘language’

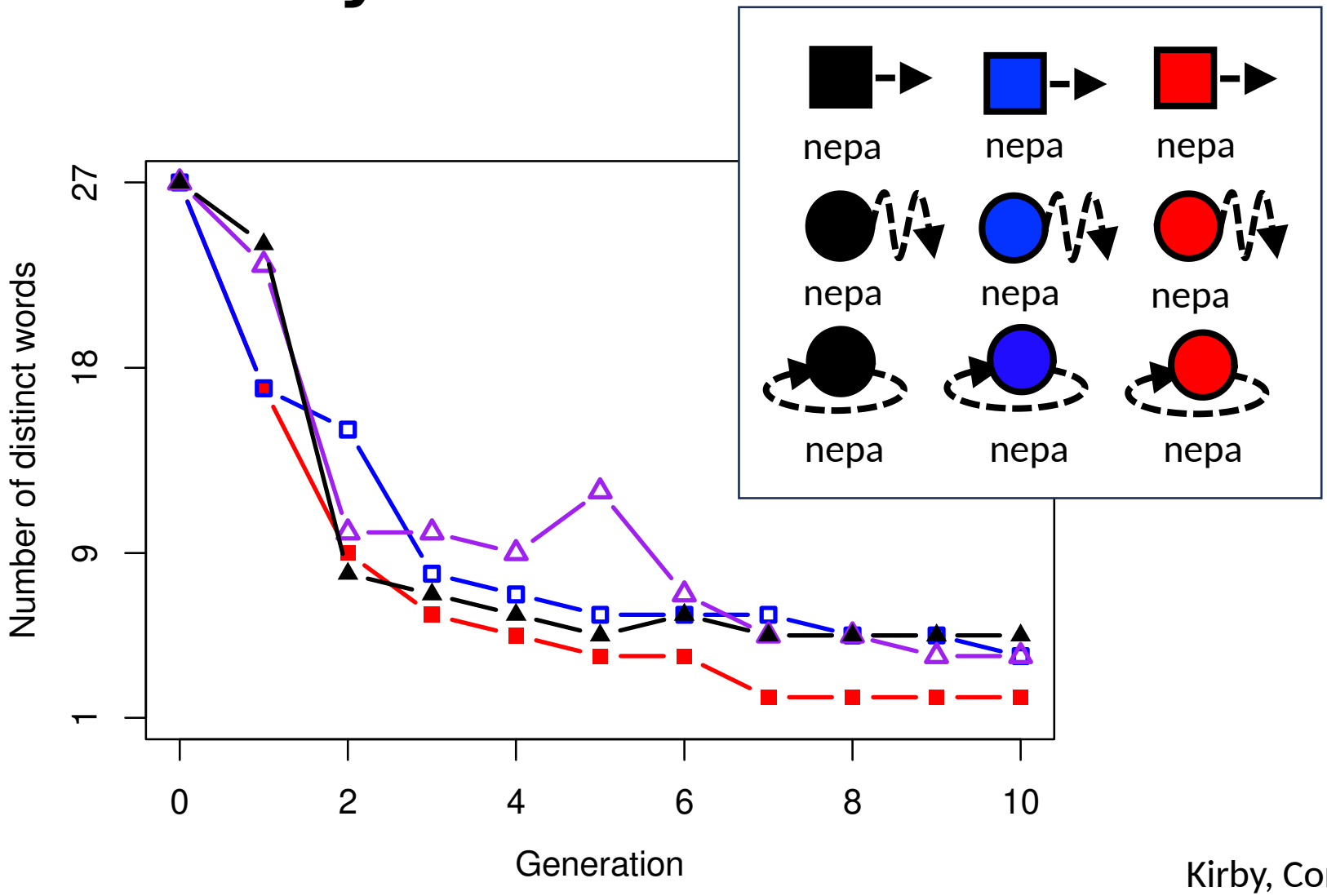


Result: loss of distinctions, degeneracy

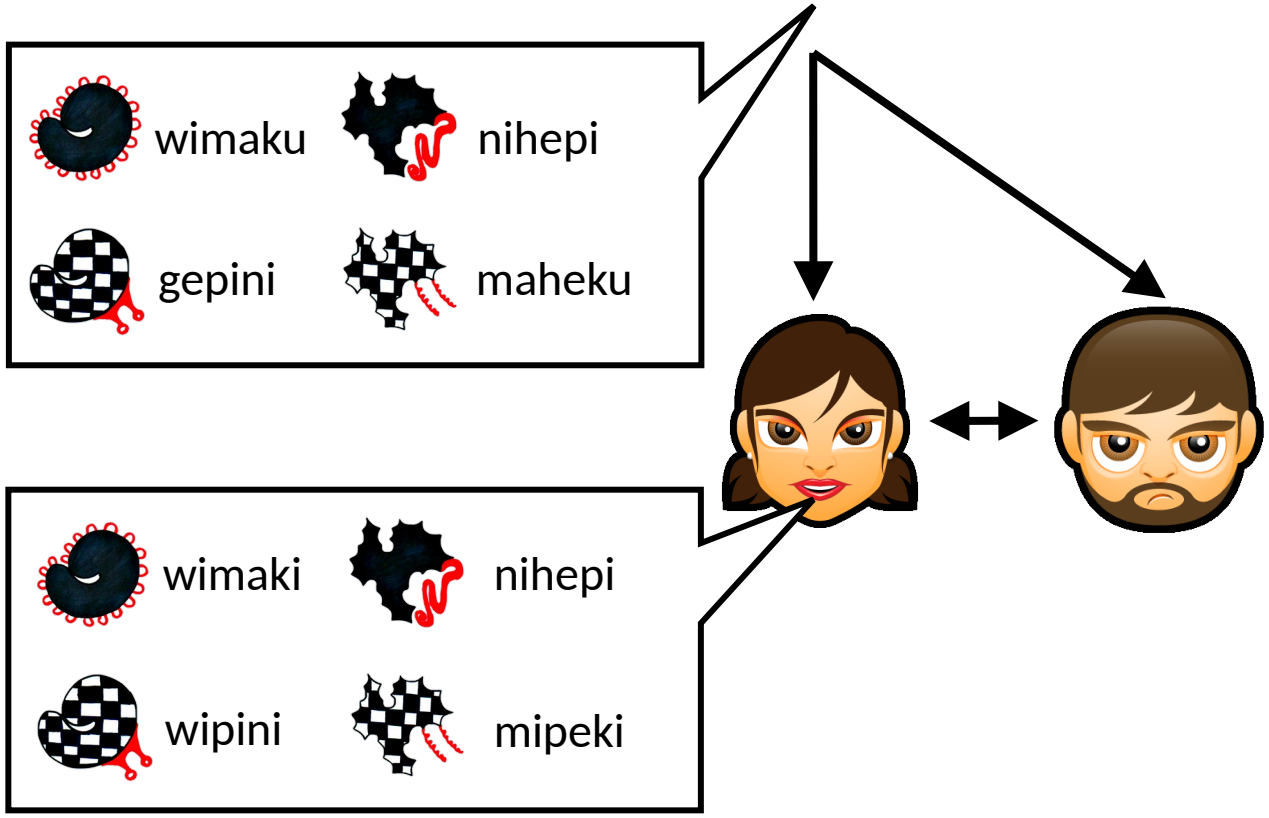


Kirby, Cornish & Smith, 2008, *PNAS*

Result: loss of distinctions, degeneracy



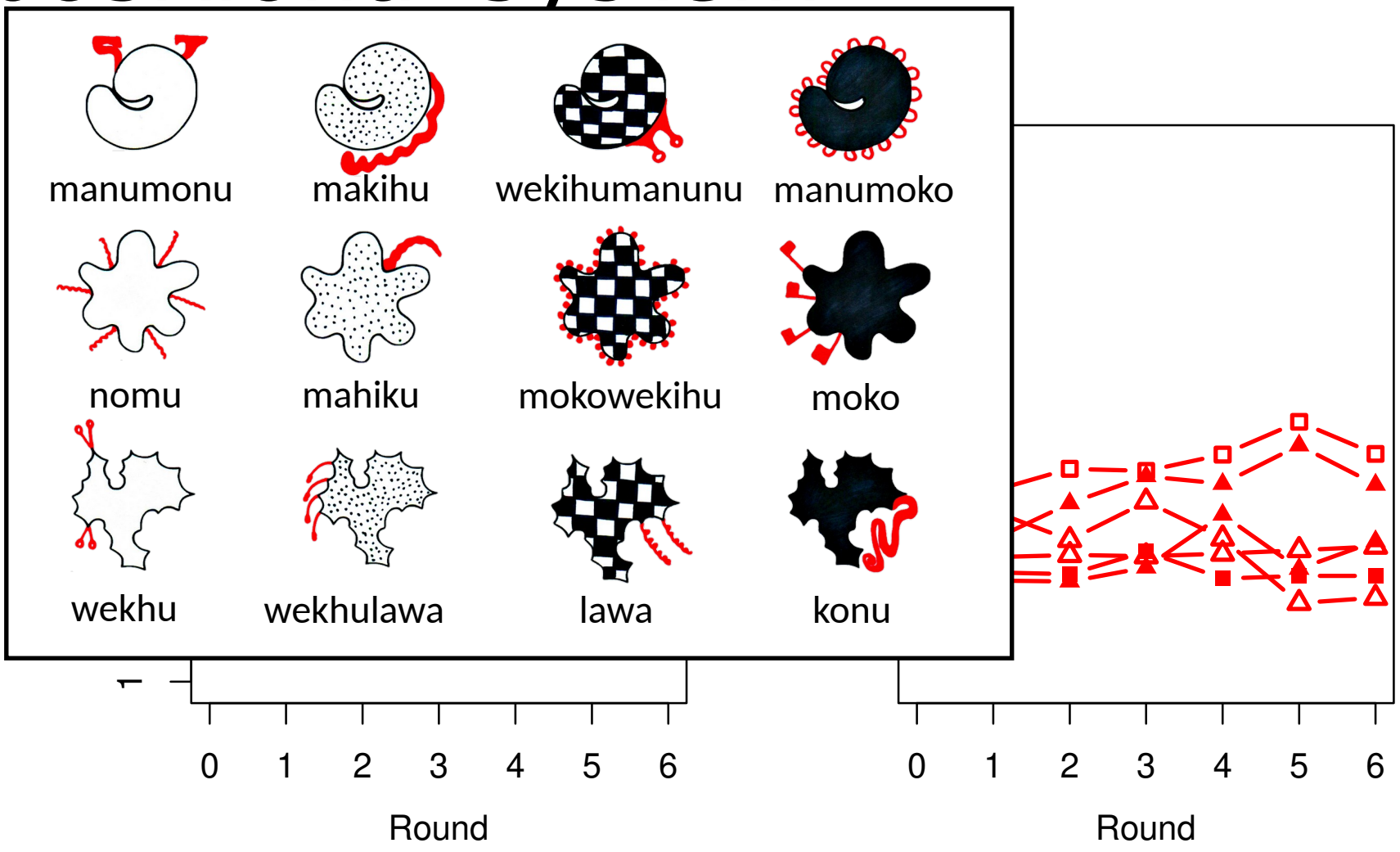
Kirby, Cornish & Smith, 2008, *PNAS*



Use only

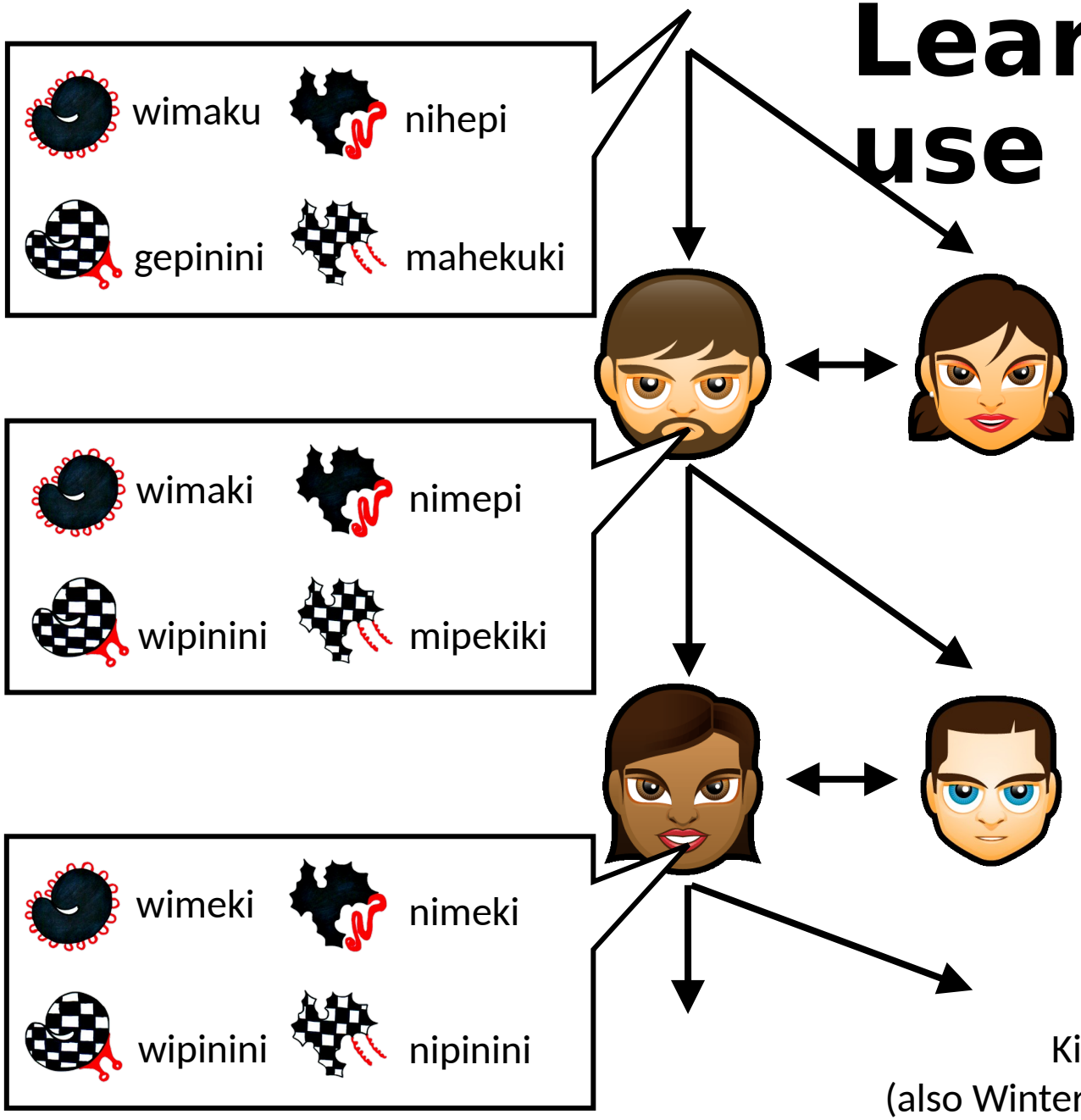
Kirby, Tamariz, Cornish & Smith, 2015, *Cognition*

Result: preservation of non-compositional system



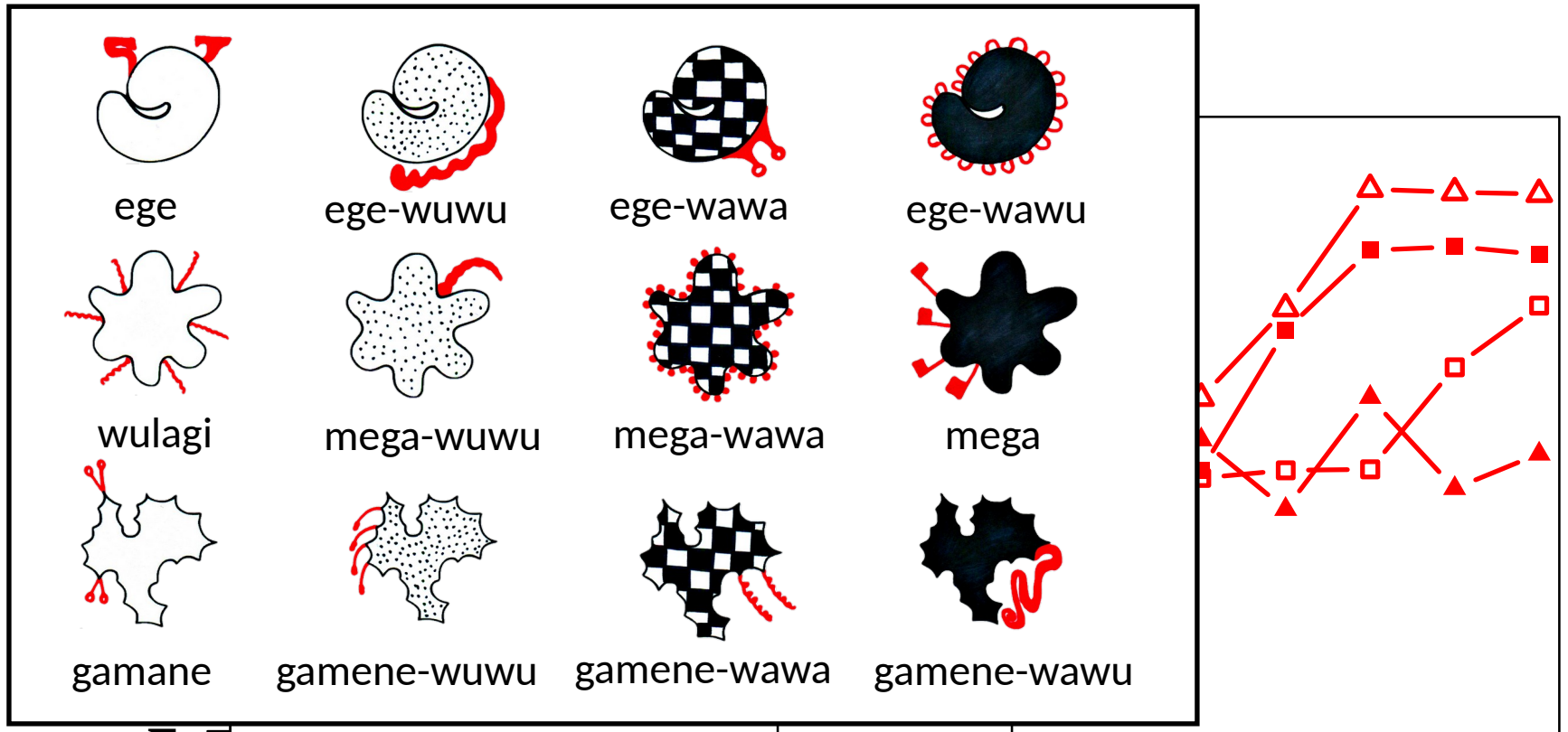
Kirby, Tamariz, Cornish & Smith, 2015, *Cognition*

Learning + use



Kirby, Tamariz, Cornish & Smith, 2015, *Cognition*
(also Winters, Kirby & Smith, 2015, *Language & Cognition*;
Saldana, Kirby, Truswell & Smith , 2019, *JoLE*)

Result: (simple) compositional structure

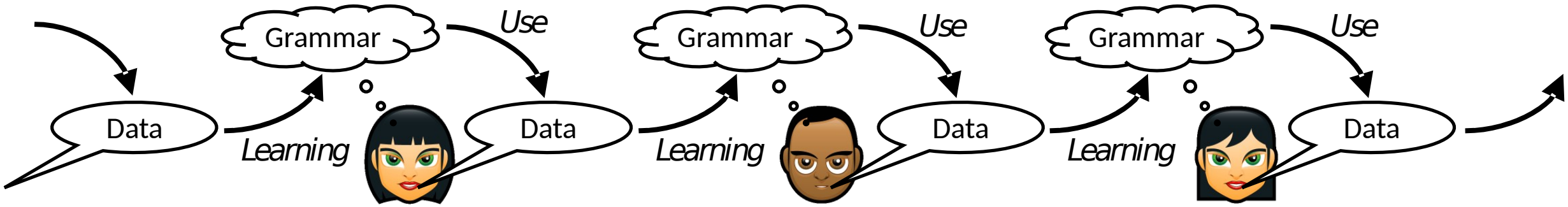


Kirby, Tamariz, Cornish & Smith, 2015, *Cognition*

Summary of Part 1

Learning plus use leads to regular compositional structure

- Learning favours simplicity
- Communicative use prevents collapse to degeneracy
- To be both learnable and useful, be regular and compositional



Part 2: same processes explain where irregularity appears **within** languages

Clem Ashton

Helen Sims-Williams

Smith, Ashton, & Sims-Williams, 2023, *Proc Cog Sci*
https://escholarship.org/uc/item/1mz1q97f


A Russian woman was sentenced to 21 years for trying to kill her doppelganger with a poisoned cheesecake!!!
?
?

But not all of language is regular and compositional

time, year, people, way, man, day, thing, child, work, life, …

But not all of language is regular and compositional

time, year, people, way, man, day, thing, child, work, life, …

is-was, has-had, does-did, says-said, gets-got, makes-made, goes-went, sees-saw, knows-knew, takes-took, thinks-thought, comes-came, gives-gave, **looks-looked**, …

But not all of language is regular and compositional

time, year, people, way, man, day, thing, child, work, life, …

is-was, has-had, does-did, says-said, gets-got, makes-made, goes-went, sees-saw, knows-knew, takes-took, thinks-thought, comes-came, gives-gave, **looks-looked**, …

beat around the bush, let the cat out of the bag, bite the bullet, pull their leg, spill the beans, go cold turkey, ring a bell, give them a run for their money, find their feet

But not all of language is regular and compositional

time, year, people, way, man, day, thing, child, work, life, …

is-was, has-had, does-did, says-said, gets-got, makes-made, goes-went, sees-saw, knows-knew, takes-took, thinks-thought, comes-came, gives-gave, **looks-looked**, …

beat around the bush, let the cat out of the bag, bite the bullet, pull their leg, spill the beans, go cold turkey, ring a bell, give them a run for their money, find their feet, your bum's out the window, …

Proposed mechanisms for the frequency-irregularity correlation

Learning-based: frequent items are more **resistant** to pressures for **regularity** operating in learning (e.g. analogy)

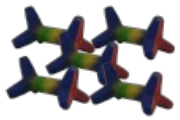
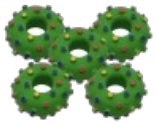
- e.g. Bybee (1995); Sims-Williams (2022)

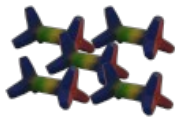
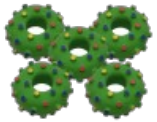
Usage-based: frequent items more **susceptible** to reductive sound change which causes **irregularity**

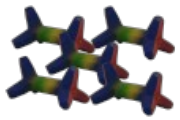
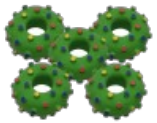
- e.g. Garrett (2015); Bybee (2017); Todd, Pierrehumbert, & Hay (2019)

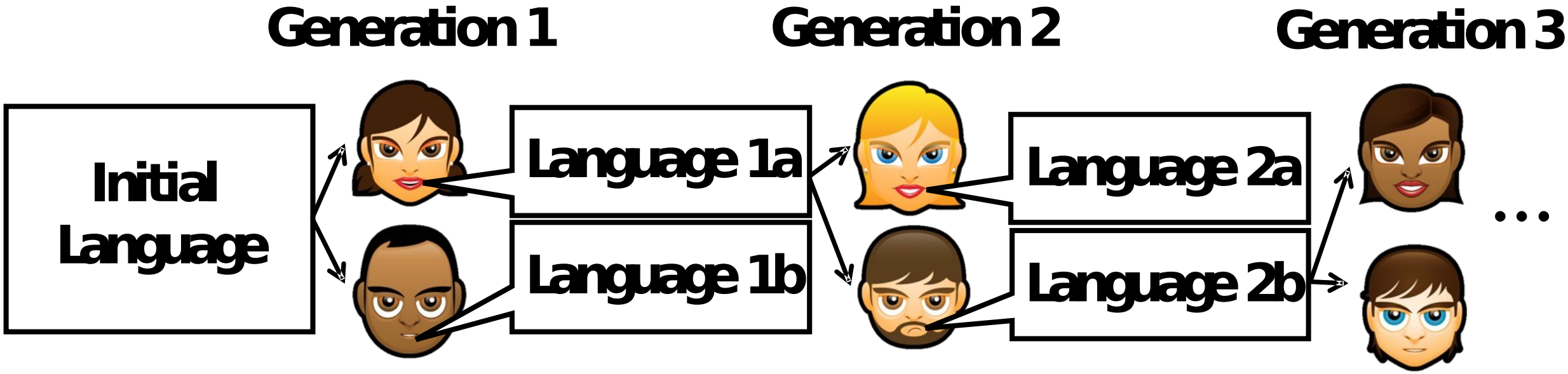
Can we reproduce the frequency-irregularity correlation under controlled conditions in the lab?

Can we test mechanisms involved?

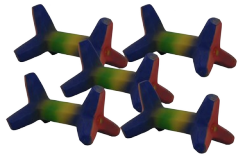
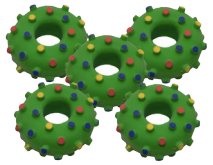
	Object 1	Object 2	Object 3	Object 4	Object 5	Object 6
Singular image						
Singular label	viza	drashru	wodra	mowo	shrunu	plonu
Plural image						
Plural label	**zawo**	drashru**bli**	**huvi**	mowo**sla**	**nuplo**	plonu**ri**

	Object 1	Object 2	Object 3	Object 4	Object 5	Object 6
Singular image						
Singular label	viza	drashru	wodra	mowo	shrunu	plonu
Plural image						
Plural label	**zawo**	drashru**bli**	**huvi**	mowo**sla**	**nuplo**	plonu**ri**
Number of trials in uniform training	16	16	16	16	16	16

	Object 1	Object 2	Object 3	Object 4	Object 5	Object 6
Singular image						
Singular label	viza	drashru	wodra	mowo	shrunu	plonu
Plural image						
Plural label	**zawo**	drashru**bli**	**huvi**	mowo**sla**	**nuplo**	plonu**ri**
Number of trials in uniform training	16	16	16	16	16	16
Number of trials in skewed training	24	24	12	12	12	12

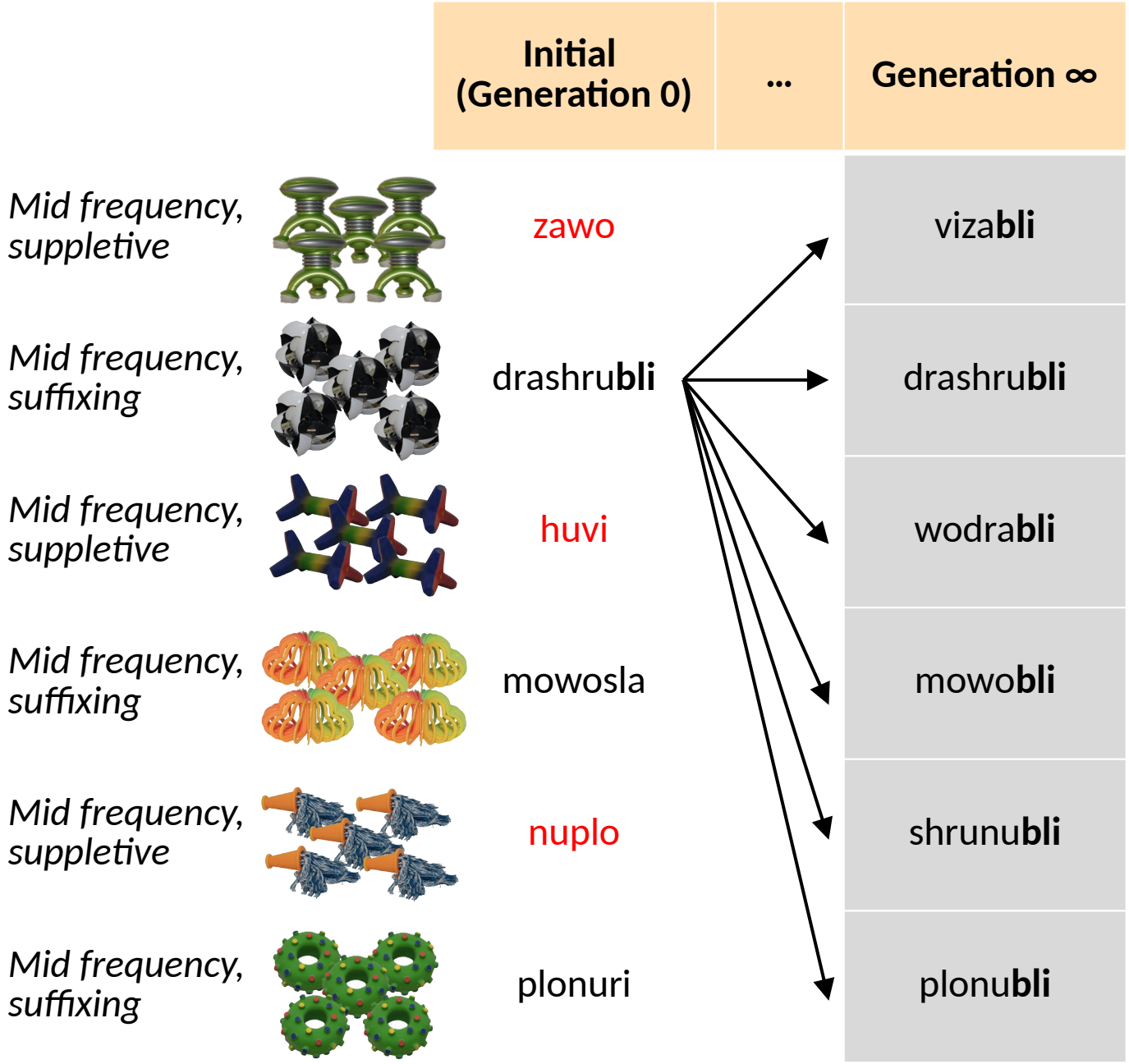

Generation 1
Generation 2
Generation 3
Initial Language
Language 1a
Language 1b
Language 2a
Language 2b
…

Prediction for uniform condition

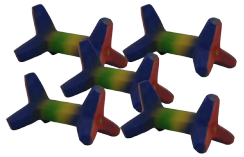
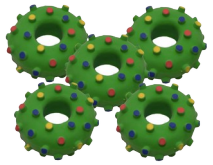
	Initial (Generation 0)	…	Generation ∞
Mid frequency, suppletive			zawo
Mid frequency, suffixing			drashru**bli**
Mid frequency, suppletive			huvi
Mid frequency, suffixing			mowosla
Mid frequency, suppletive			nuplo
Mid frequency, suffixing			plonuri

Prediction for uniform condition

Emergence of a single regular rule

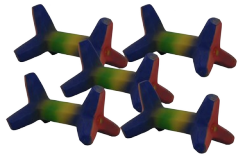
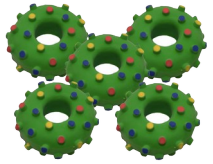


Prediction for skewed condition

	Initial (Generation 0)	…	Generation ∞
High frequency, suppletive			zawo
High frequency, suffixing			drashru**bli**
Low frequency, suppletive			huvi
Low frequency, suffixing			mowosla
Low frequency, suppletive			nuplo
Low frequency, suffixing			plonuri

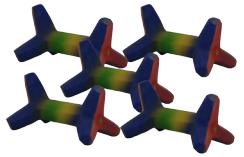
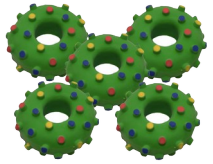
Prediction for skewed condition

A single regular rule??

	Initial (Generation 0)	...	Generation ∞
High frequency, suppletive	 zawo		viza**bli**
High frequency, suffixing	 drashru**bli**		drashru**bli**
Low frequency, suppletive	 huvi		wodra**bli**
Low frequency, suffixing	 mowosla		mowo**bli**
Low frequency, suppletive	 nuplo		shrunu**bli**
Low frequency, suffixing	 plonuri		plonu**bli**

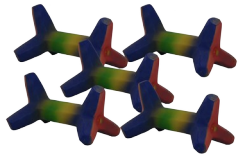
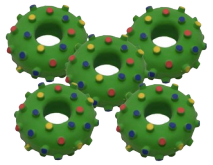
Prediction for skewed condition

High frequency items remain exceptional?

	Initial (Generation 0)	...	Generation ∞
High frequency, suppletive 	zawo	→	zawo
High frequency, suffixing 	drashrubli	→	drashrubli
Low frequency, suppletive 	huvi	→	wodra**sla**
Low frequency, suffixing 	mowo**sla**	→	mowo**sla**
Low frequency, suppletive 	nuplo	→	shrunu**sla**
Low frequency, suffixing 	plonuri	→	plonu**sla**

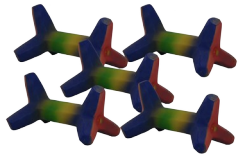
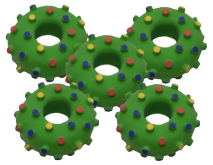
Prediction for skewed condition

High frequency suppletive remains exceptional, high frequency suffix becomes the regular?

	Initial (Generation 0)	...	Generation ∞
High frequency, suppletive	 zawo	→	zawo
High frequency, suffixing	 drashru**bli**	→	drashru**bli**
Low frequency, suppletive	 huvi	→	wodra**bli**
Low frequency, suffixing	 mowosla	→	mowo**bli**
Low frequency, suppletive	 nuplo	→	shrunu**bli**
Low frequency, suffixing	 plonuri	→	plonu**bli**

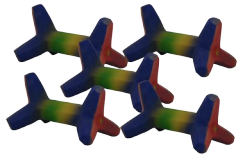
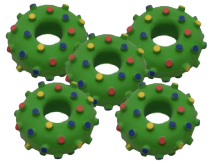
Prediction for skewed condition

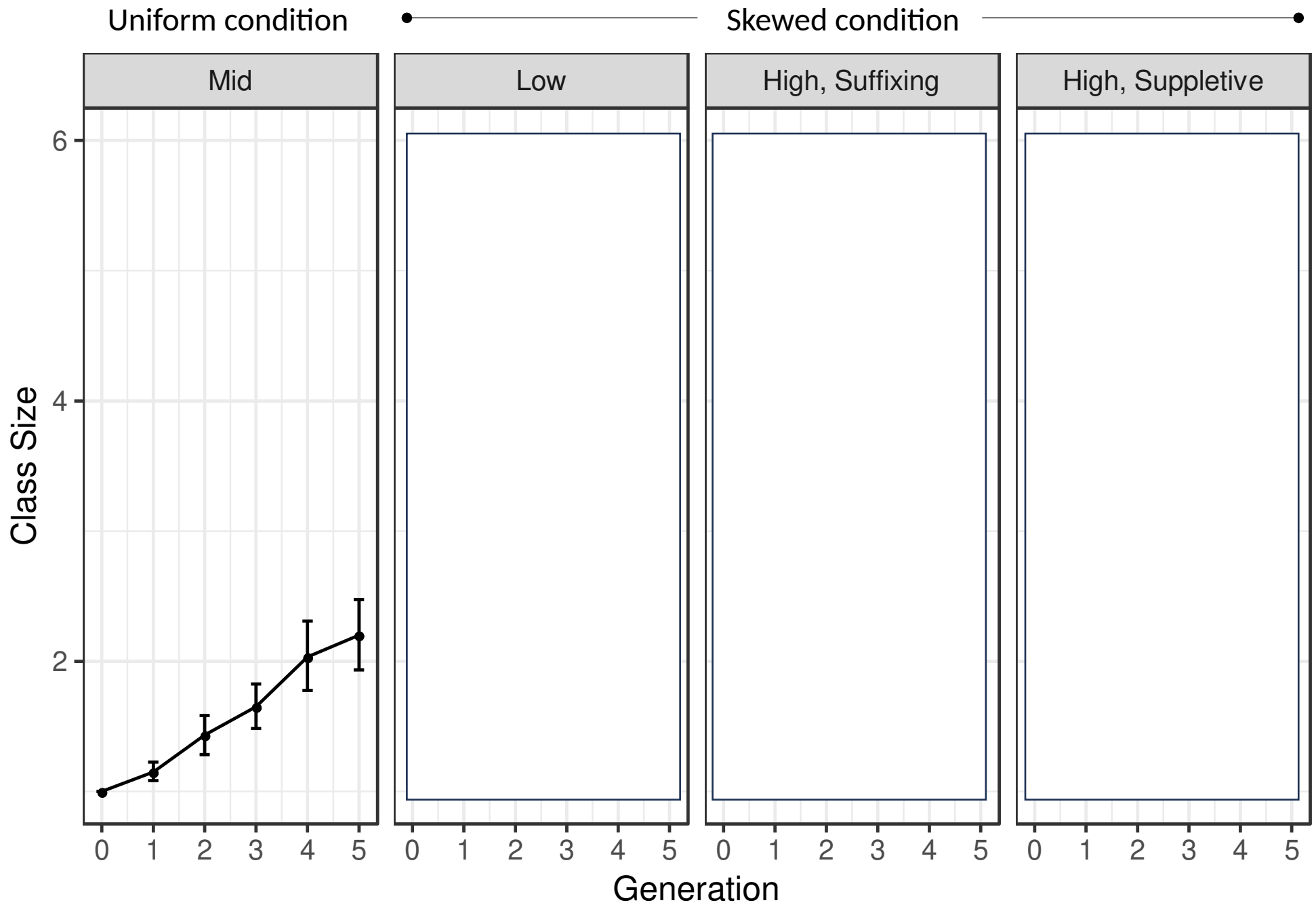
High frequency suppletive remains exceptional, high frequency suffix becomes the regular?

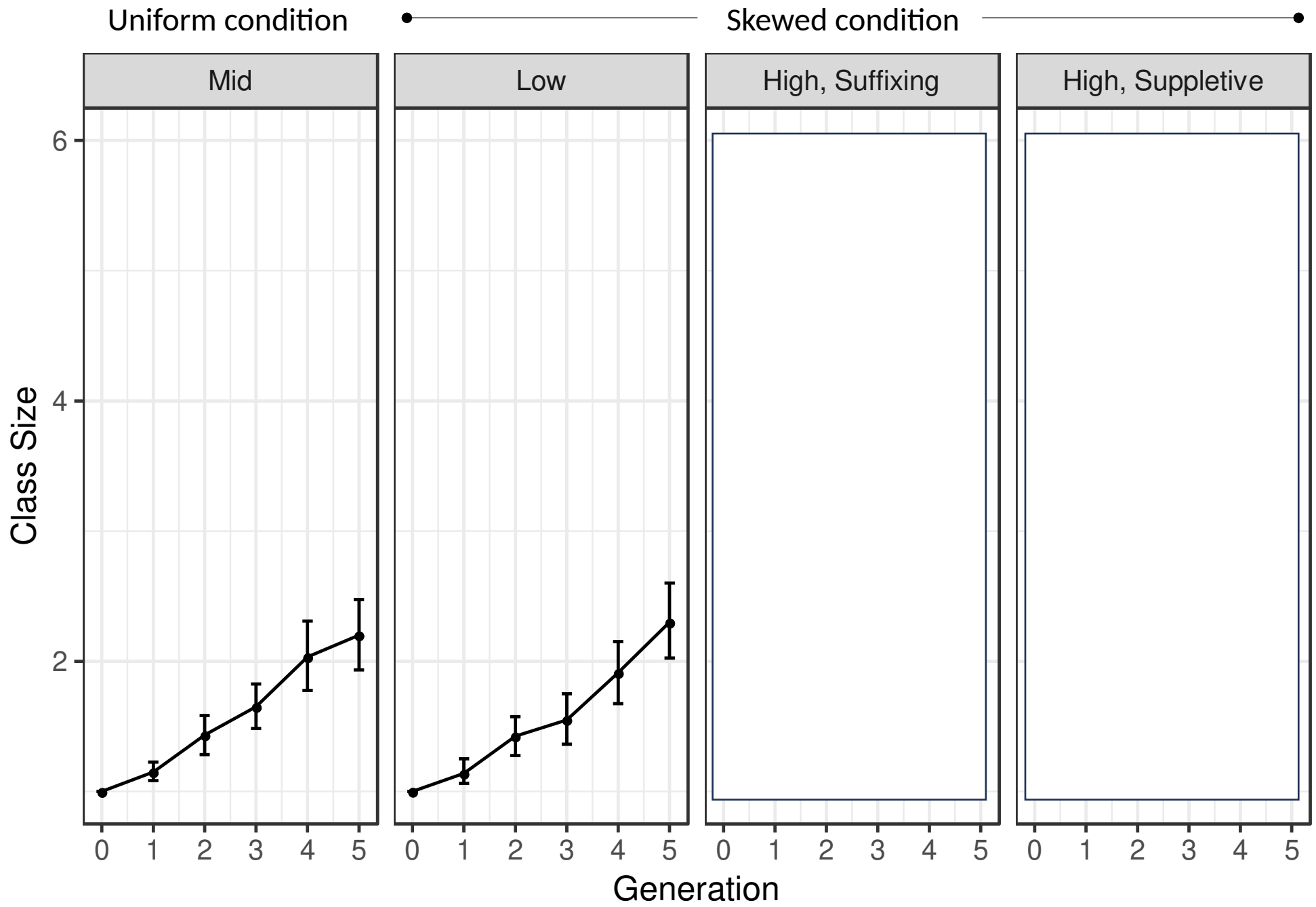
	Initial (Generation 0)		...	Generation ∞
High frequency, suppletive		zawo	1 →	zawo 1
High frequency, suffixing		drashru**bli**	1 →	drashru**bli** 5
Low frequency, suppletive		huvi	1 →	wodra**bli** 5
Low frequency, suffixing		mowosla	1 →	mowo**bli** 5
Low frequency, suppletive		nuplo	1 →	shrunu**bli** 5
Low frequency, suffixing		plonuri	1 →	plonu**bli** 5

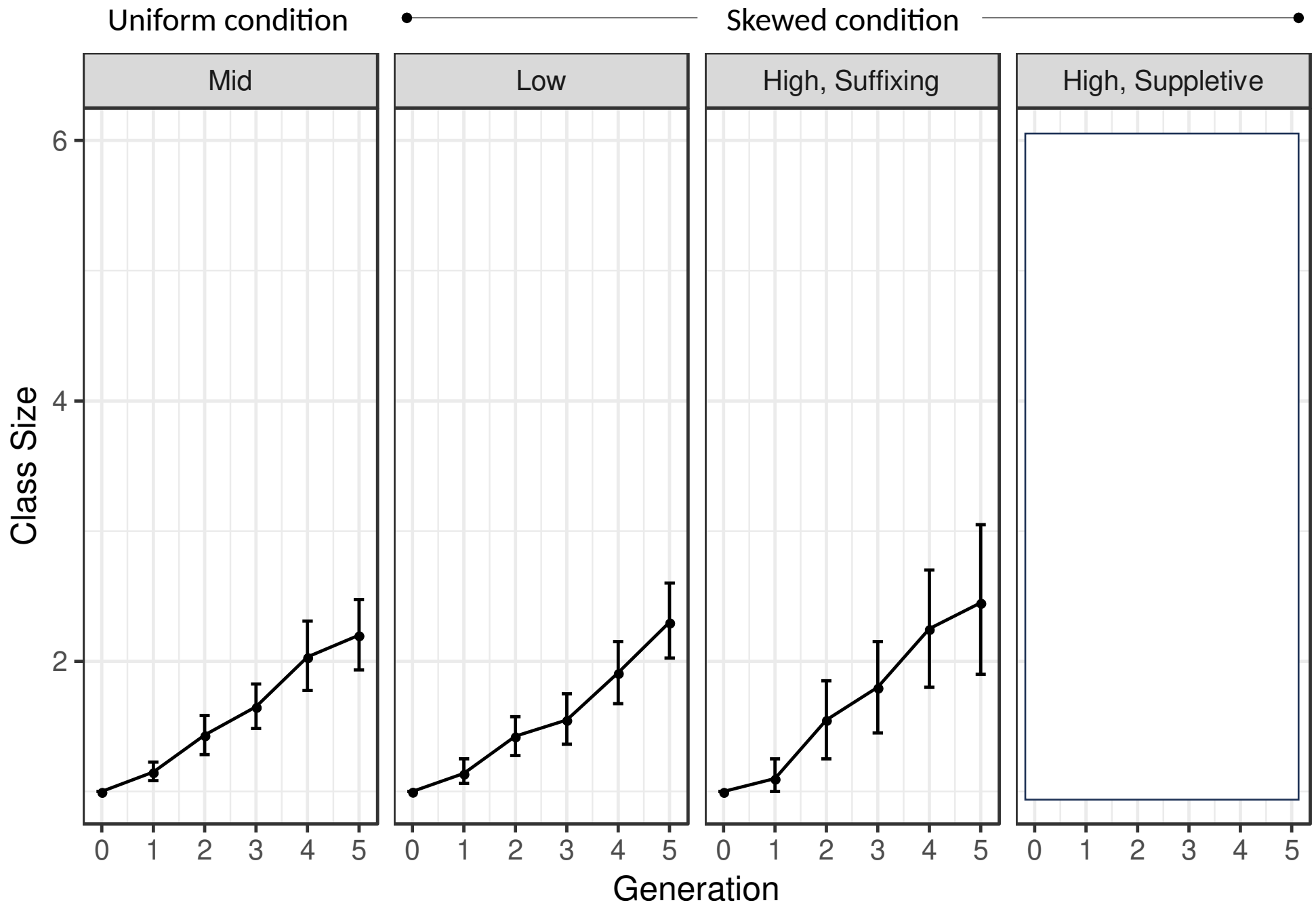
Prediction for skewed condition

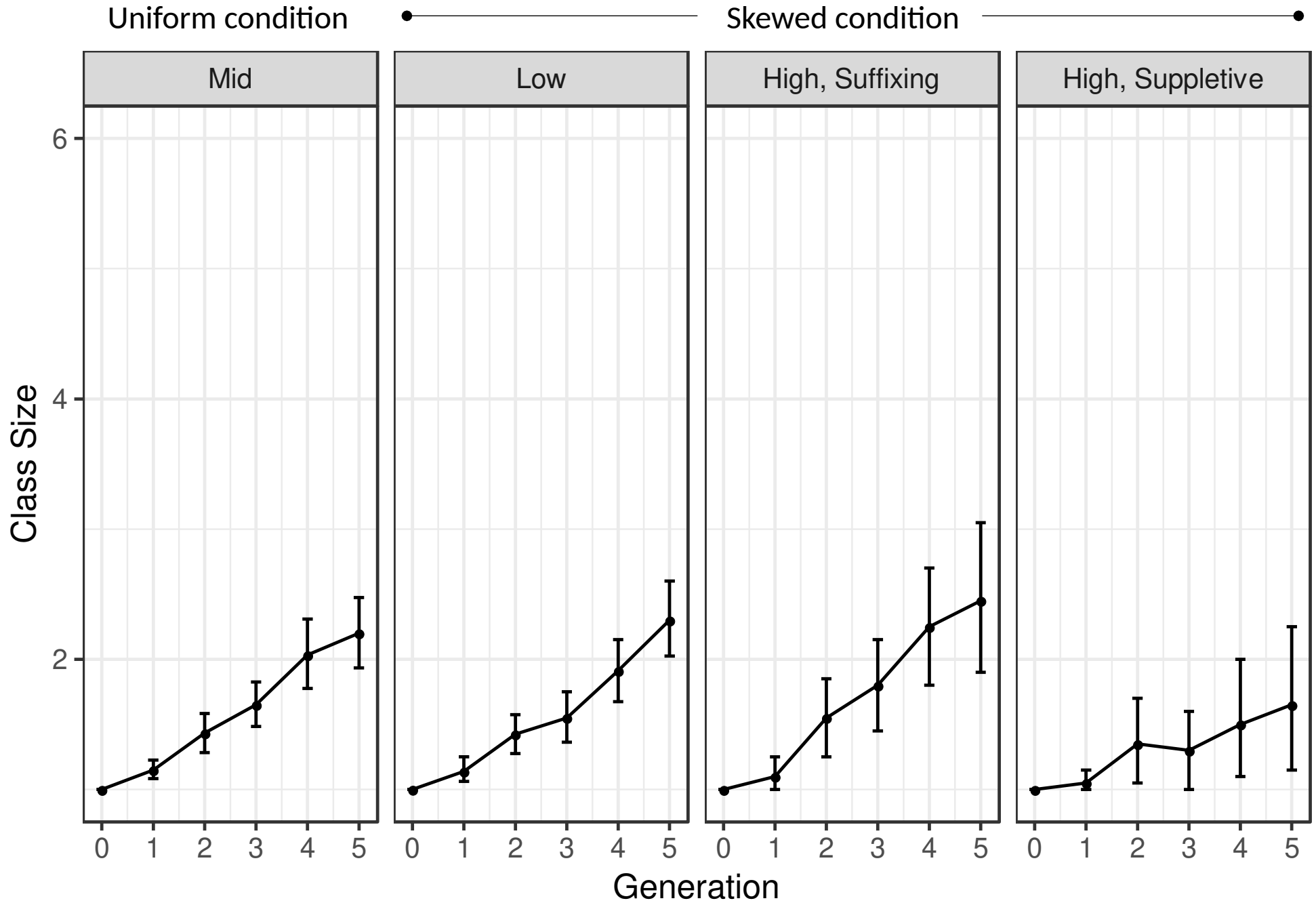
A single regular rule??

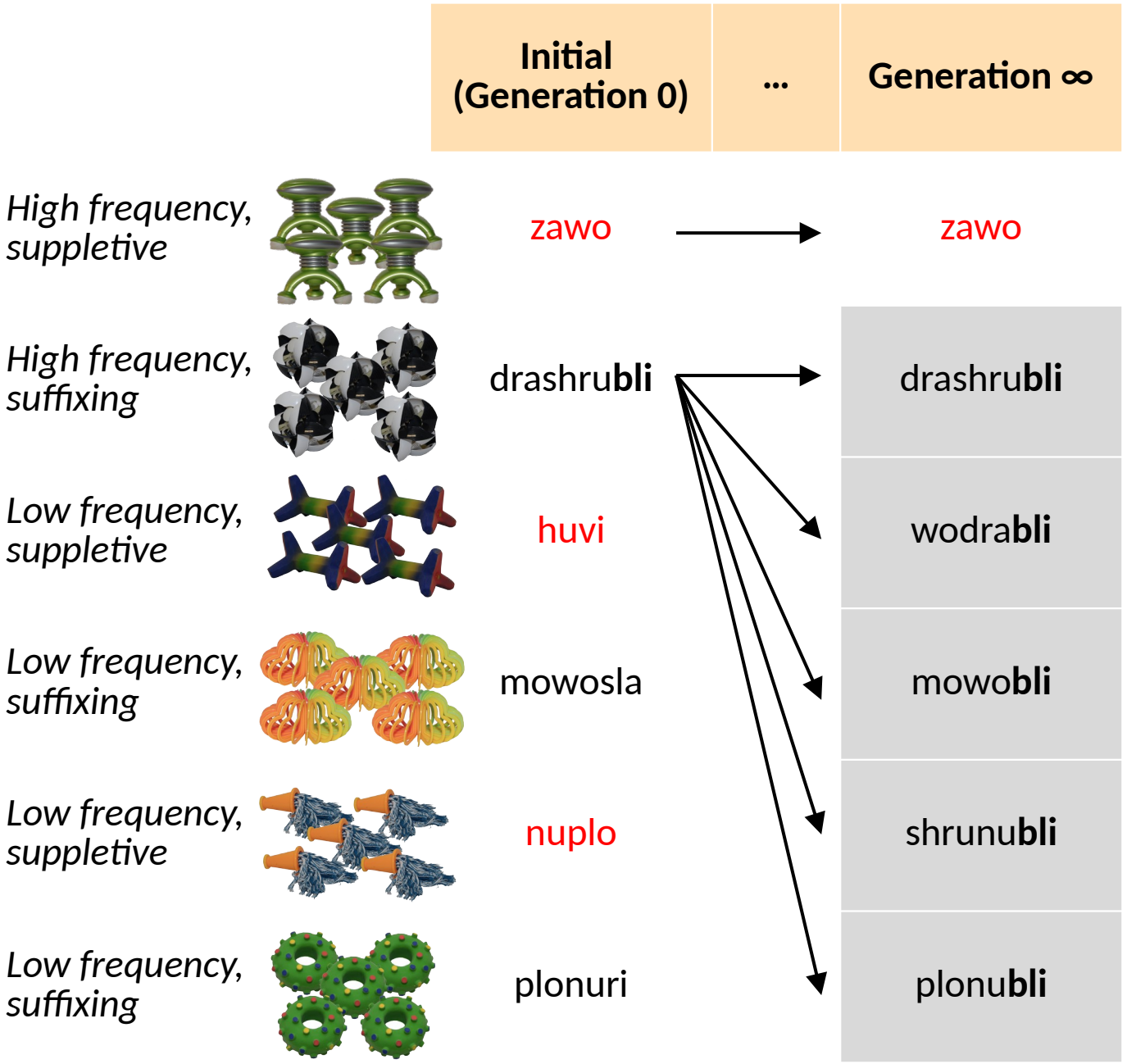
	Initial (Generation 0)		...	Generation ∞	
High frequency, suppletive		zawo	1	vizabli	6
High frequency, suffixing		drashrubli	1	drashrubli	6
Low frequency, suppletive		huvi	1	wodrabli	6
Low frequency, suffixing		mowosla	1	mowobli	6
Low frequency, suppletive		nuplo	1	shrunubli	6
Low frequency, suffixing		plonuri	1	plonubli	6


Uniform condition
Skewed condition
Mid
Low
High, Suffixing
High, Suppletive
Class Size
6
4
2
0
1
2
3
4
5
Generation


Uniform condition
Skewed condition
Mid
Low
High, Suffixing
High, Suppletive
Class Size
6
4
2
0
1
2
3
4
5
Generation


Uniform condition
Skewed condition
Mid
Low
High, Suffixing
High, Suppletive
Class Size
6
4
2
0
1
2
3
4
5
Generation


Uniform condition
Skewed condition
Mid
Low
High, Suffixing
High, Suppletive
Class Size
6
4
2
0
1
2
3
4
5
Generation



In the Skewed condition

High-frequency suppletives **resist** regularization

High-frequency suffixes **become the regular**, attracting lower-frequency items to their class

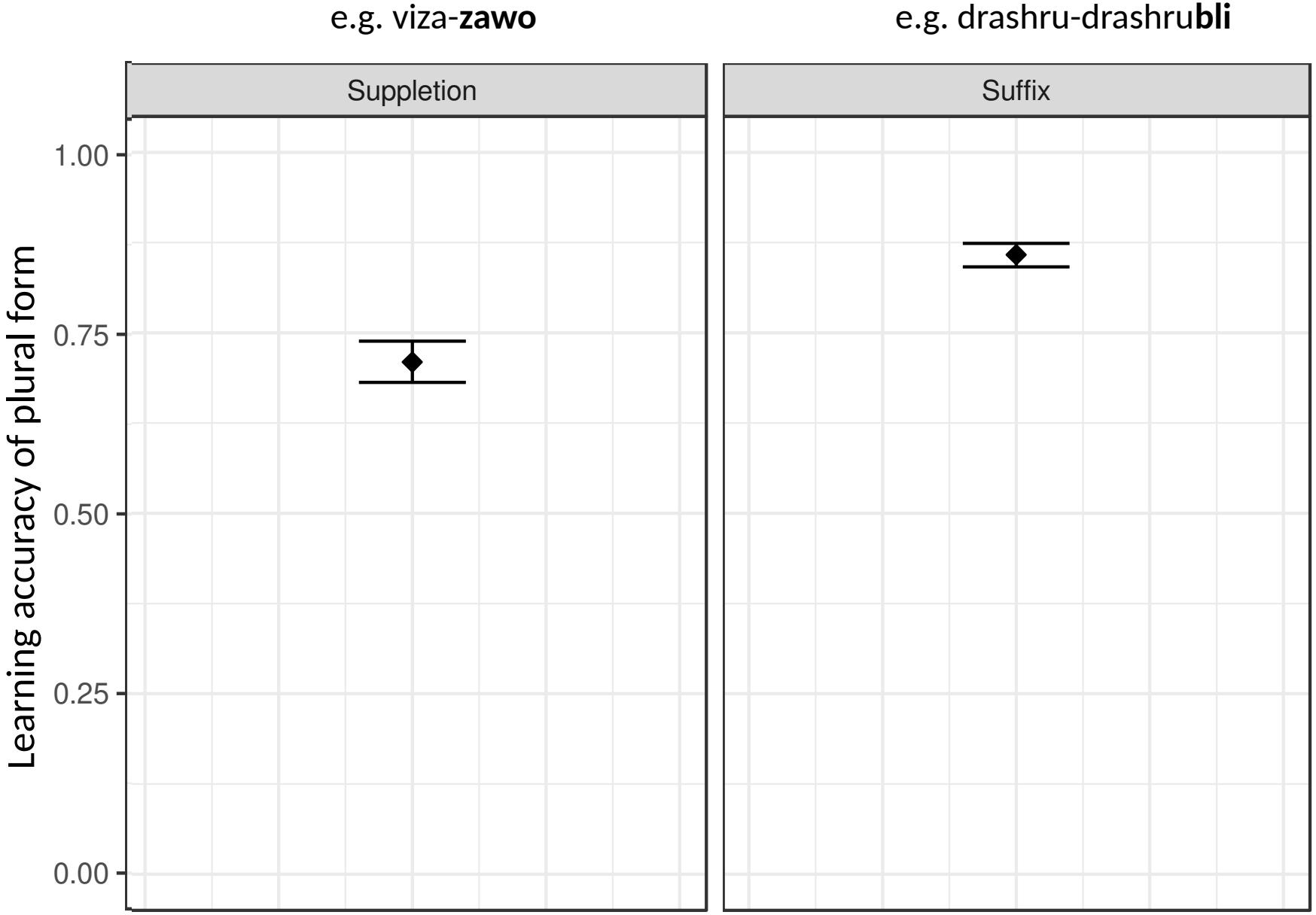
Reminder: Proposed mechanisms for the frequency-irregularity correlation

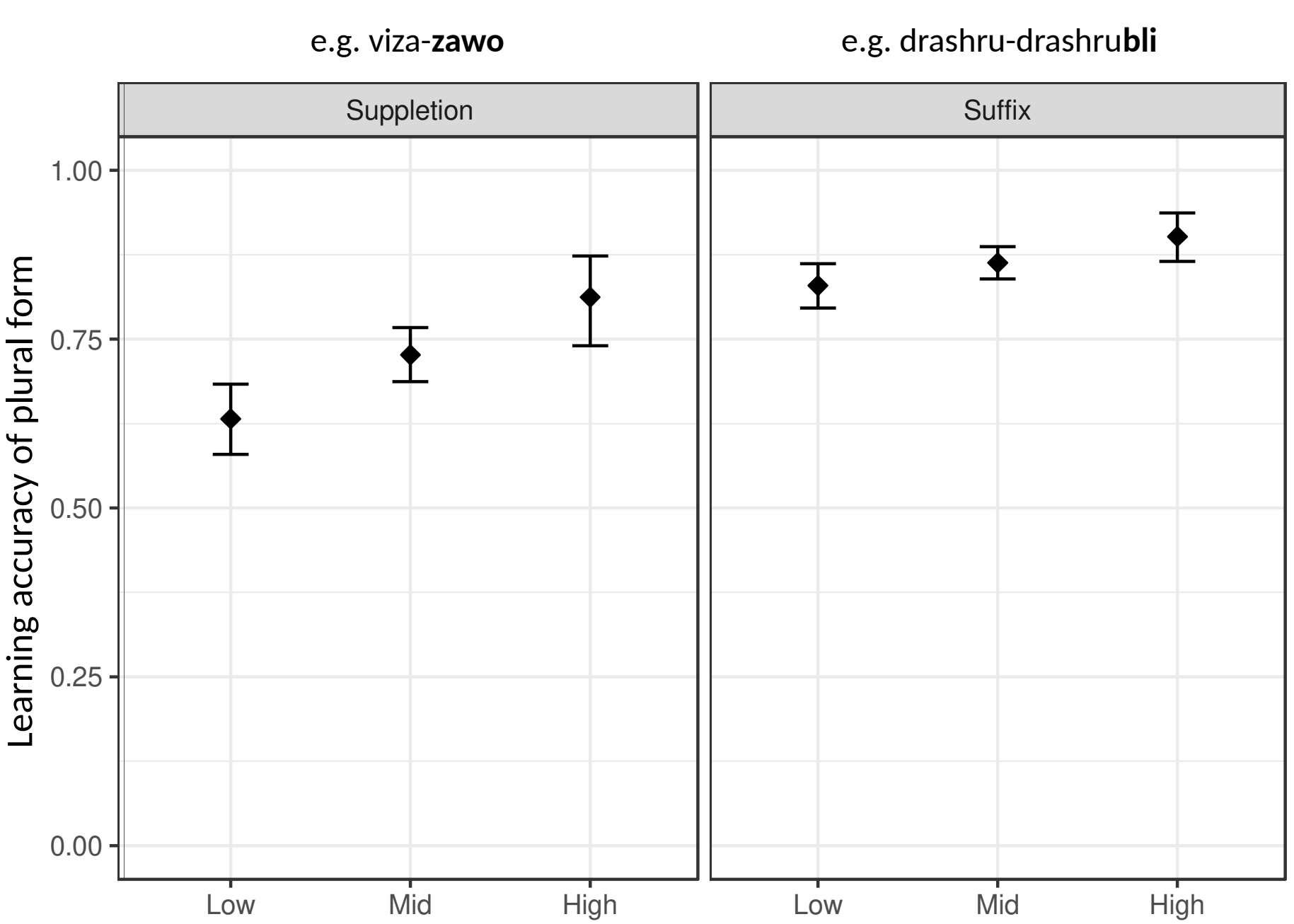
Learning-based: frequent items are more **resistant** to pressures for **regularity** operating in learning (e.g. analogy)

- e.g. Bybee (1995); Sims-Williams (2022)

Usage-based: frequent items more **susceptible** to reductive sound change which causes **irregularity**

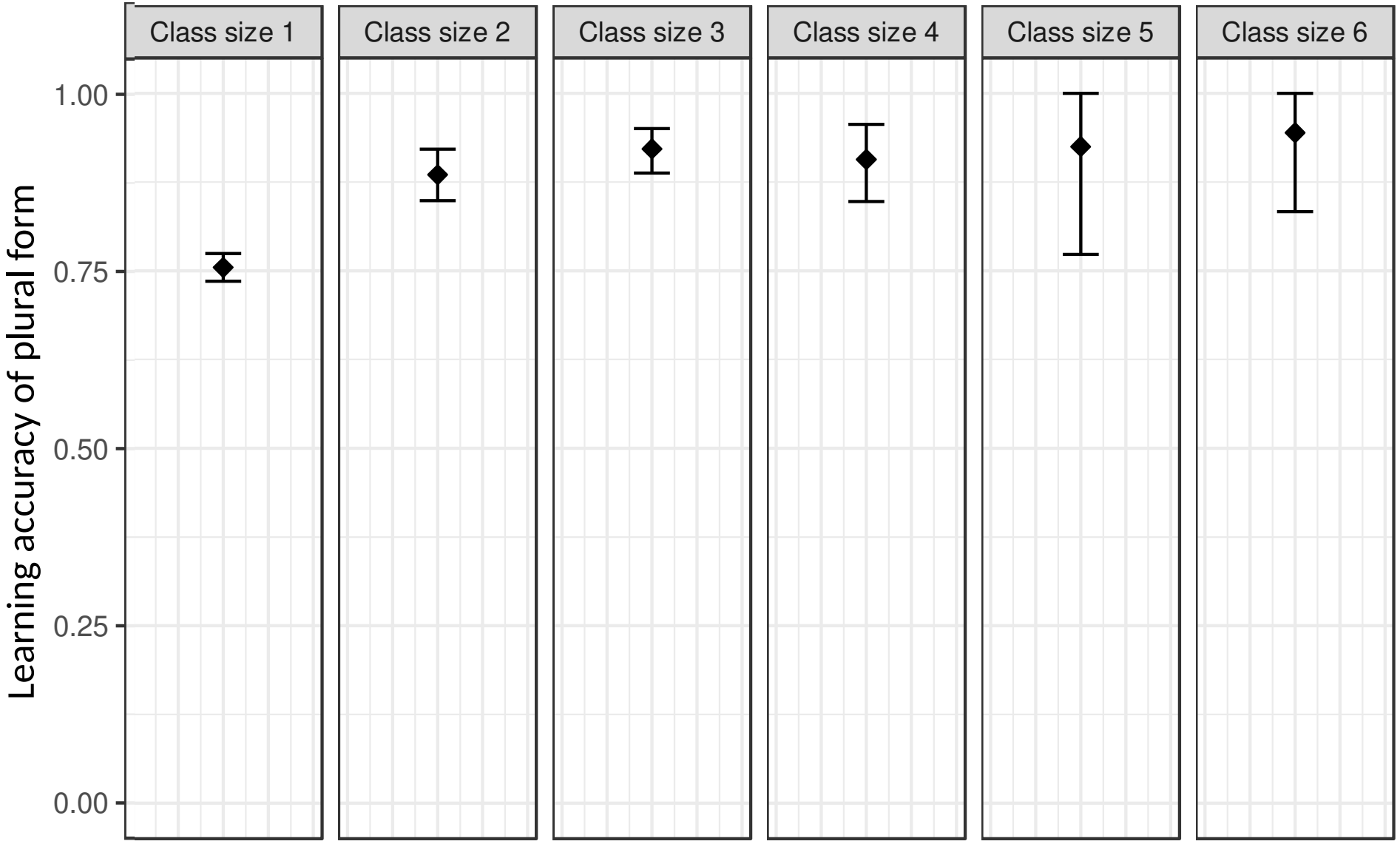
- e.g. Garrett (2015); Bybee (2017); Todd, Pierrehumbert, & Hay (2019)


e.g. viza-**zawo**
e.g. drashru-drashru**bli**
Suppletion
Suffix
Learning accuracy of plural form
1.00
0.75
0.50
0.25
0.00


e.g. viza-**zawo**
e.g. drashru-drashru**bli**
Suppletion
Suffix
Learning accuracy of plural form
1.00
0.75
0.50
0.25
0.00
Low
Mid
High
Low
Mid
High

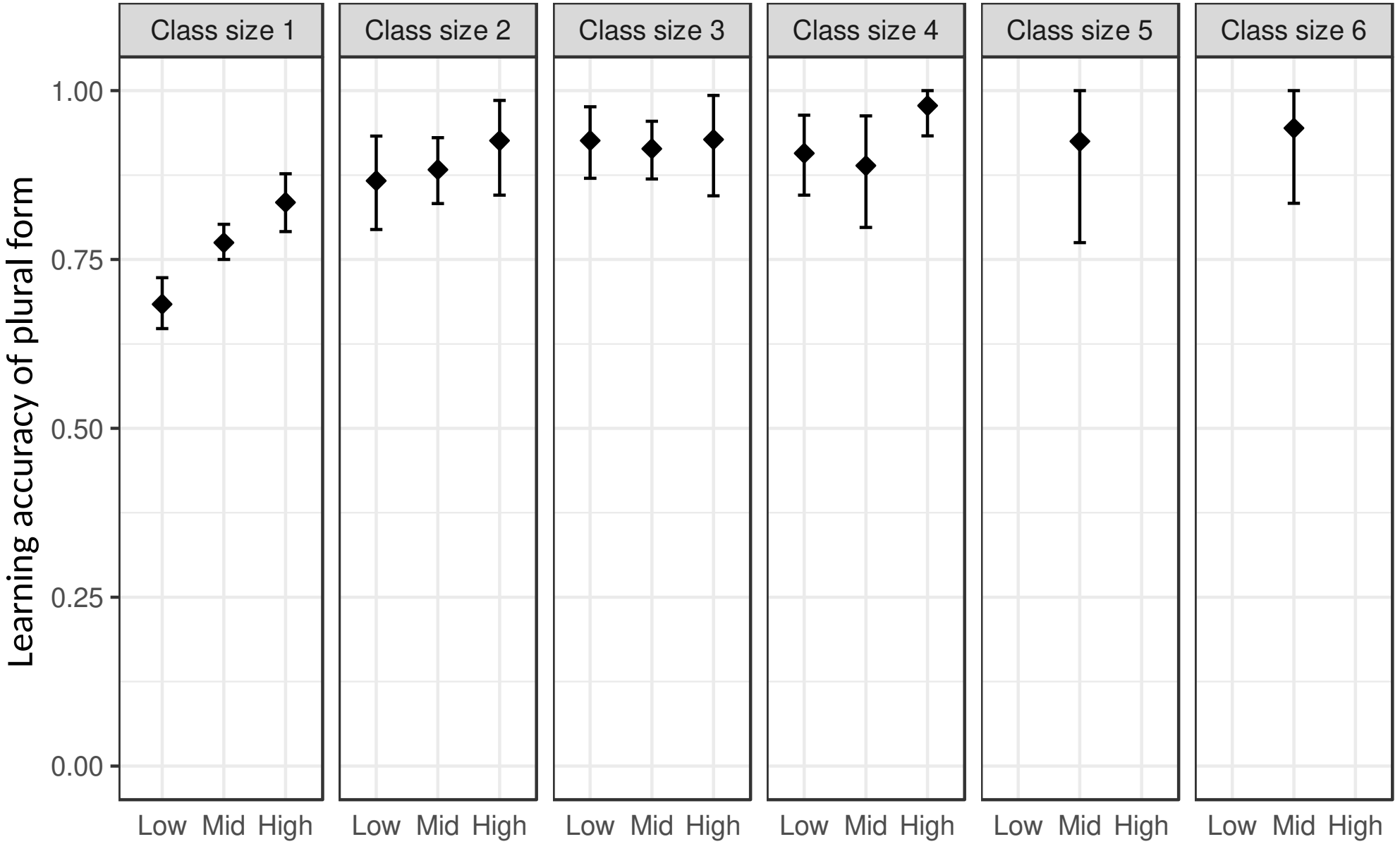
e.g. viza-**zawo**
e.g. drashru-drashru**bli**

e.g. drashru-drashru**bli**, viza-viza**bli**,
wodra-wodra**bli**, shrunu-shrunu**bli**, …



e.g. viza-**zawo**
e.g. drashru-drashru**bli**

e.g. drashru-drashru**bli**, viza-viza**bli**,
wodra-wodra**bli**, shrunu-shrunu**bli**, ...



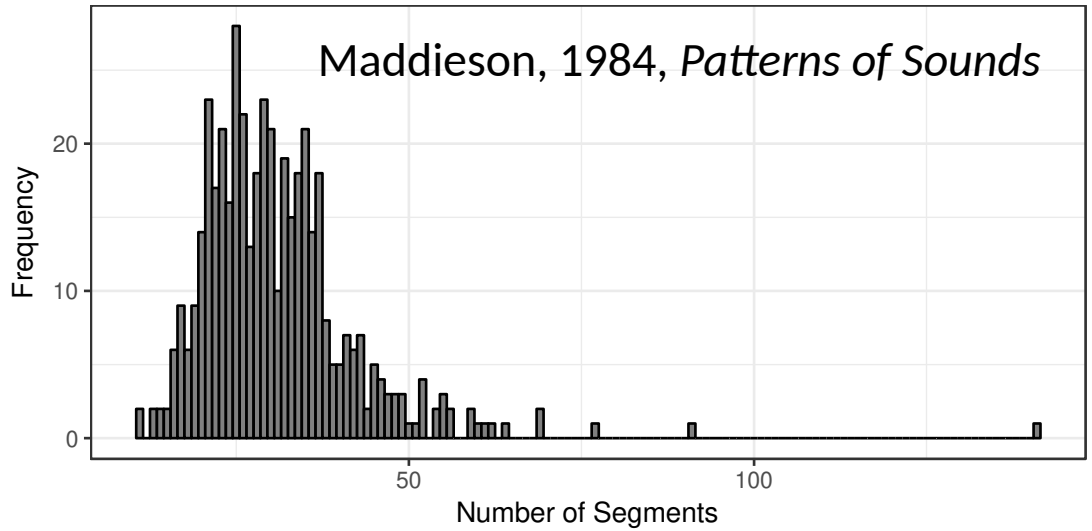
Summary of Part 2

- Things regularize (as we'd expect based on Part 1)
- High frequency items behave differently
  - High-frequency suppletives **resist** regularization, yielding a frequency-irregularity correlation
  - High-frequency suffixes **become the regular**
- High frequency provides (some) protection from the learning penalty associated with being idiosyncratic, and provides a niche where irregularity can persist
- In relation to part 1: learning pressure depends on frequency and therefore differs in different parts of the grammar

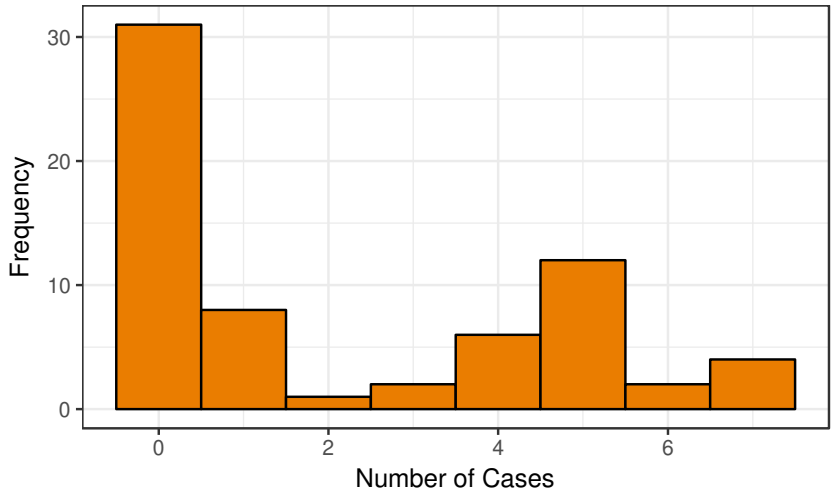
Part 3: same processes also explain where irregularity appears **across** languages

Smith, 2024, *Proc Cog Sci*
https://osf.io/preprints/psyarxiv/2kzd4

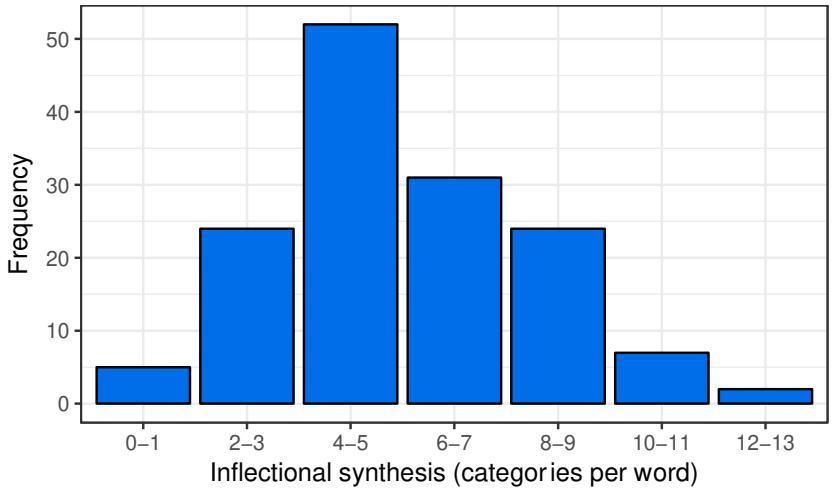
Languages differ in their complexity



Zaslavsky et al, 2019, *Cognitive Neuropsychology*

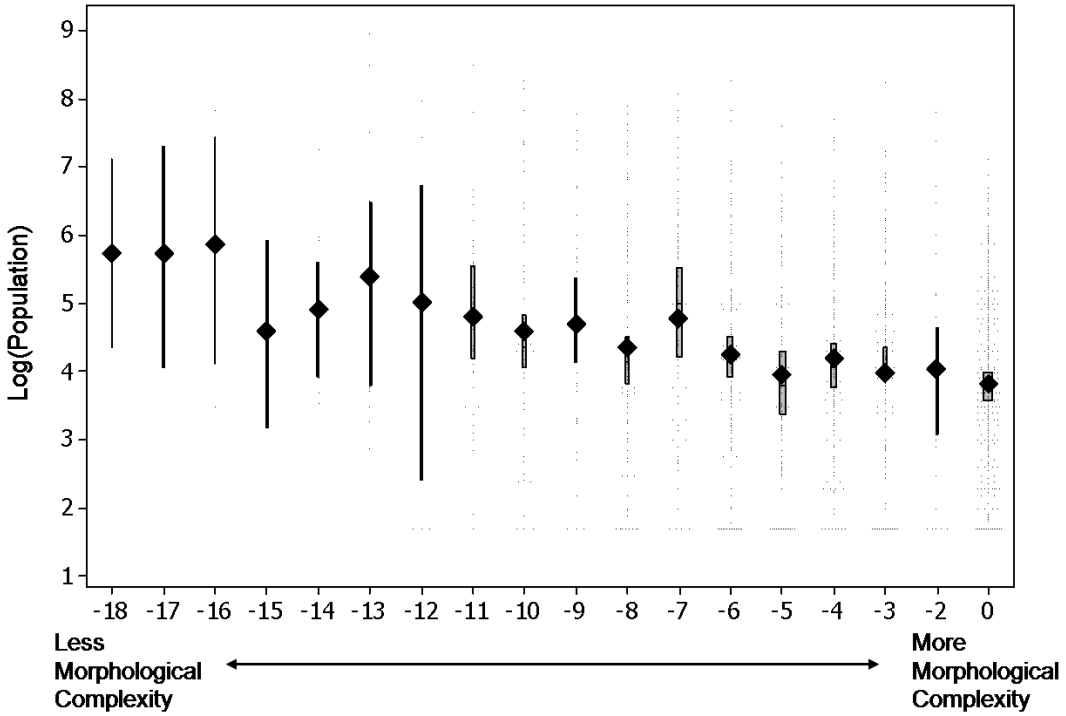


Winter & Bentz, 2013, *LDC*

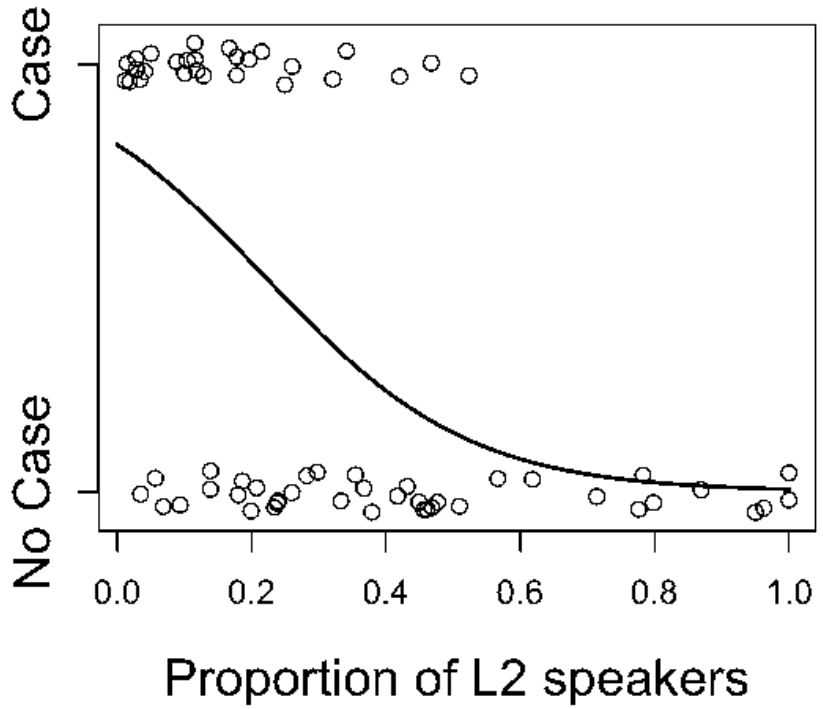


Bickel & Nichols, 2013, in *WALS*

The Linguistic Niche Hypothesis: Languages spoken in larger populations, with more non-native speakers, tend to be simpler*



Lupyan & Dale, 2010, *PLOS ONE*



Winter & Bentz, 2013, *LDC*

* Maybe: see also e.g. Koplenig, 2019, *Roy Soc Open Science;* Kauhanen, Einhaus & Walkden, 2023, *JoLE;* Koplenig, 2023, arXiv; Shcherbakova et al., 2023, *Science Advances*

Potential mechanism: uncertainty in interaction

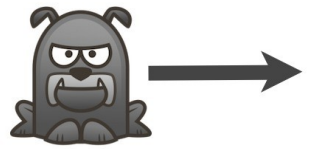
e.g. Atkinson et al., 2017, *JoLE*; Raviv et al., 2019, *Proc Roy Soc*; Feher et al., 2019, *JML*; Frank & Smith, 2020, *Proc Cog Sci*; Loy et al., 2020, *Discourse Processes*; …

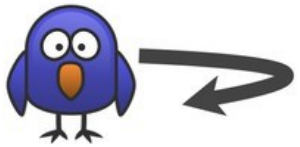
Potential mechanism: learning from heterogenous sources; **learning from non-native speakers**

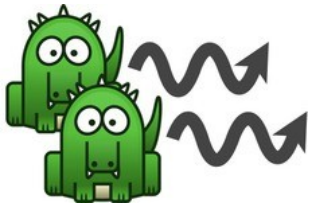
Atkinson et al., 2018, *Cognitive Science*; Berdicevskis & Semenuks, 2022, *PLOS ONE*; **this talk**

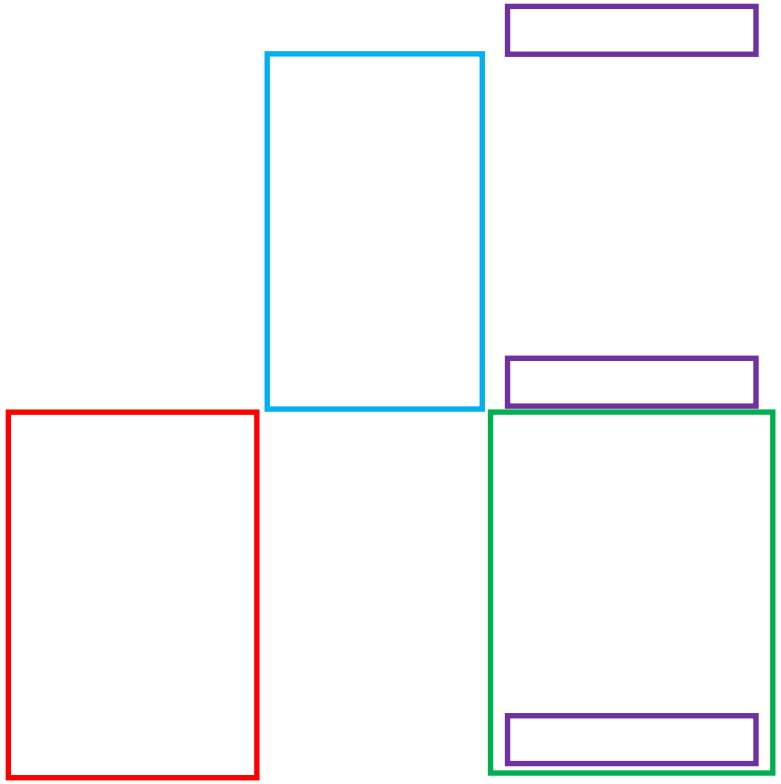
Experiment 1: establishing an experimental proxy for native vs non-native speakers

Target language





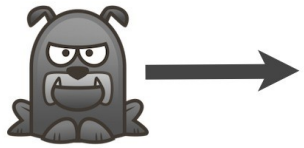




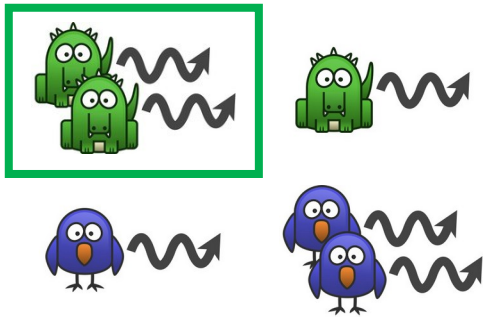
Training
(18 trials)

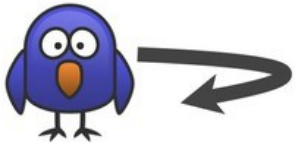
Testing
(18 trials)

wona grolo wooshan



sumuk snapop boingesp?





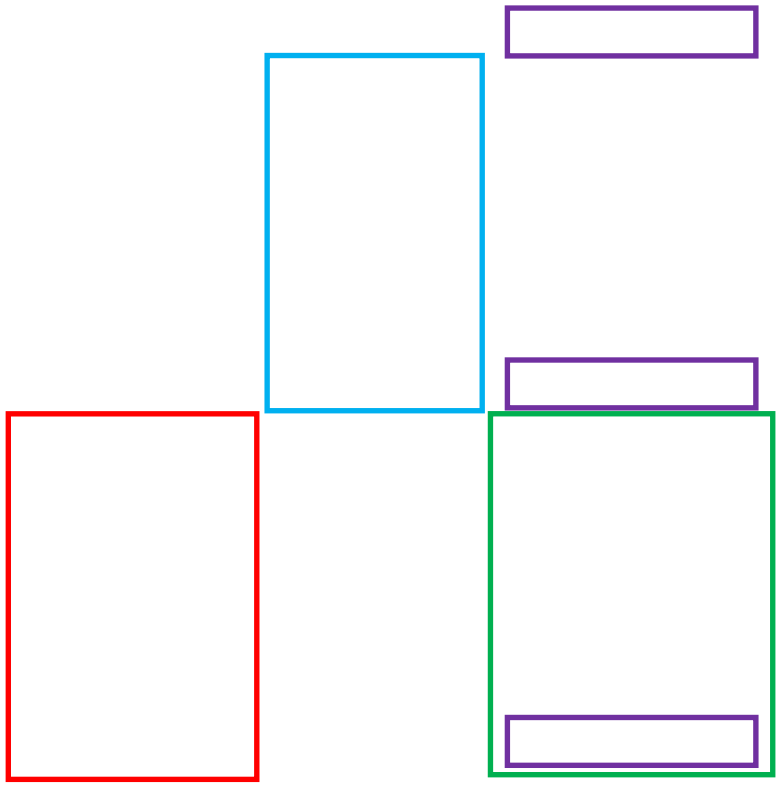
wona twito loop

Repeat
(up to 8 rounds)

depends on
number + animal

depends on
number + animal

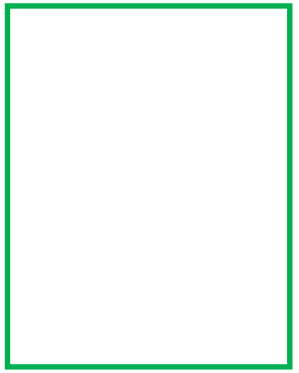
depends on
number + animal + movement



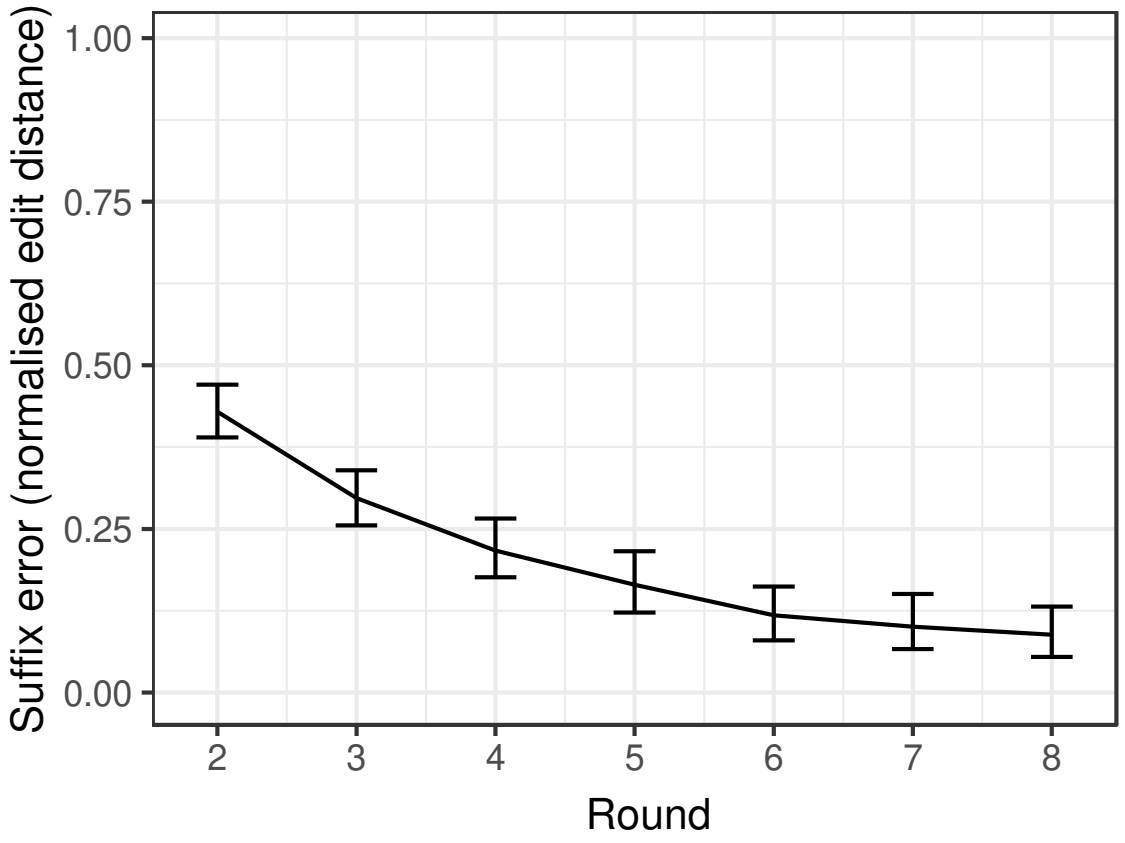
depends on ∅

depends on number

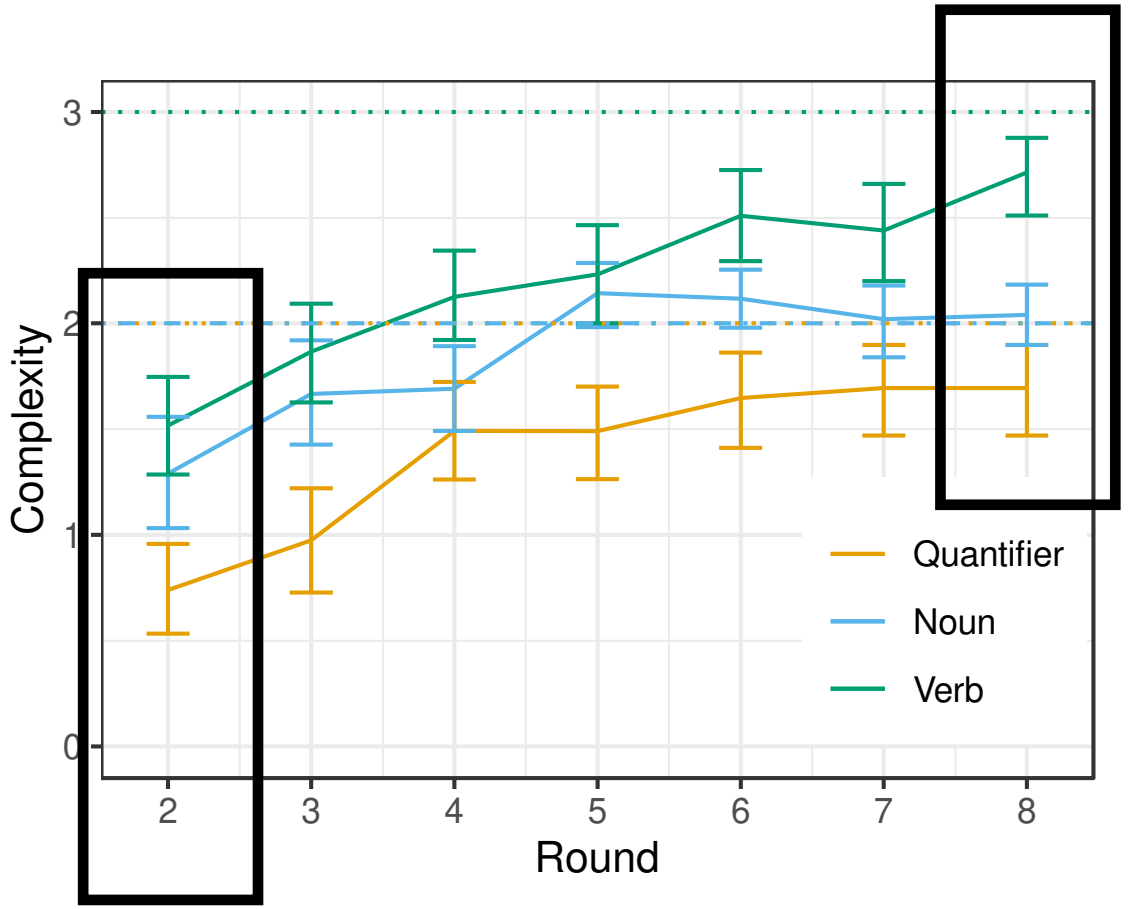
depends on number + animal



Experiment 1 results (N=94)



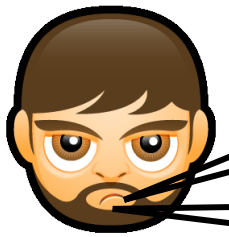
Proxy for native speakers
(≈ no simplification)



Proxy for non-native speakers
(simplification)

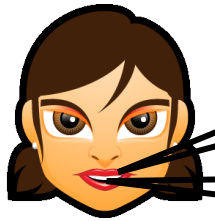
Experiment 2: investigating the effect of population size and proportion of non-native speakers on a single step of intergenerational transmission

Experiment 1 participants



Round 2
-a -o -an, -a -op -asp, -a -o -an, …

Round 8
-a -o -an, -ak -op -asp, -a -o -an, …



Round 2
-ak -o -an, -ak -op -esp, -a -o -esp, …

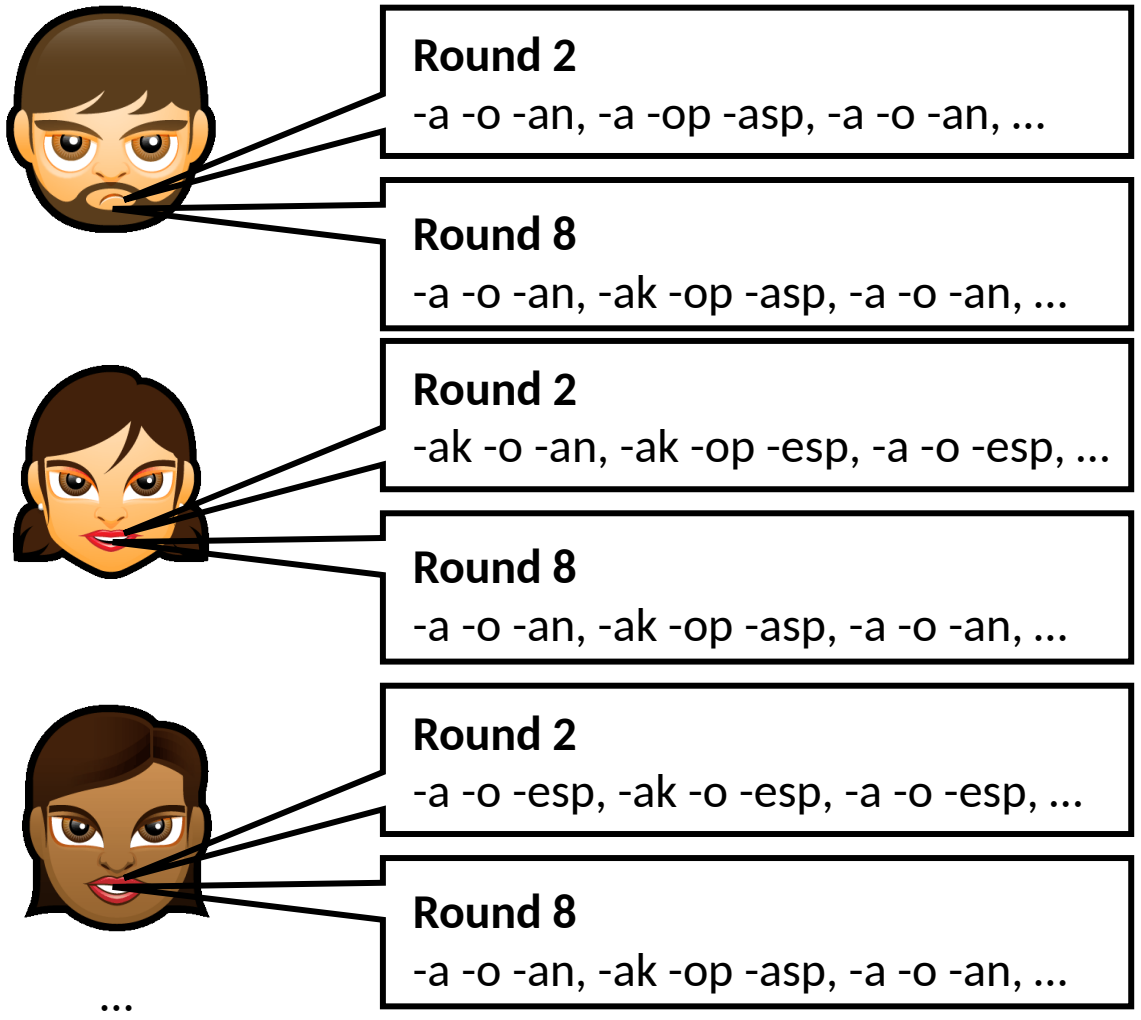
Round 8
-a -o -an, -ak -op -asp, -a -o -an, …

Round 2
-a -o -esp, -ak -o -esp, -a -o -esp, …

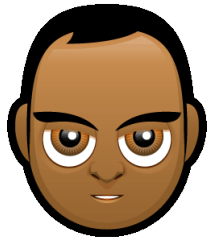
Round 8
-a -o -an, -ak -op -asp, -a -o -an, …

…

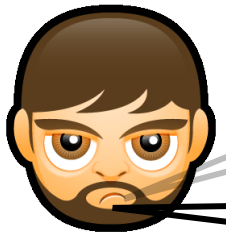
Experiment 1 participants



Experiment 2 participant

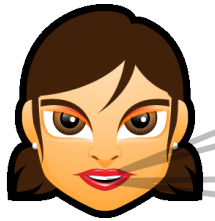


Experiment 1 participants



Round 2
-a -o -an, -a -op -asp, -a -o -an, …

Round 8
-a -o -an, -ak -op -asp, -a -o -an, …



Round 2
-ak -o -an, -ak -op -esp, -a -o -esp, …

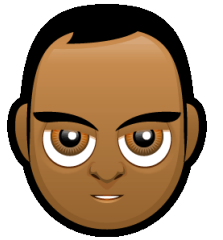
Round 8
-a -o -an, -ak -op -asp, -a -o -an, …

Round 2
-a -o -esp, -ak -o -esp, -a -o -esp, …

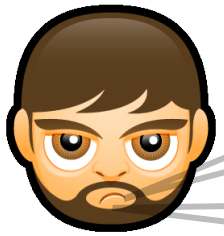
Round 8
-a -o -an, -ak -op -asp, -a -o -an, …

…

Experiment 2 participant

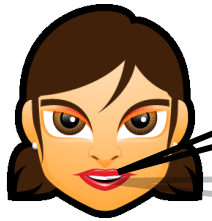


Experiment 1 participants



Round 2
-a -o -an, -a -op -asp, -a -o -an, …

Round 8
-a -o -an, -ak -op -asp, -a -o -an, …



Round 2
-ak -o -an, -ak -op -esp, -a -o -esp, …

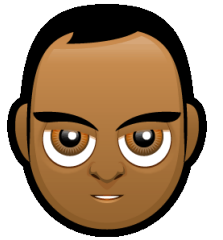
Round 8
-a -o -an, -ak -op -asp, -a -o -an, …

Round 2
-a -o -esp, -ak -o -esp, -a -o -esp, …

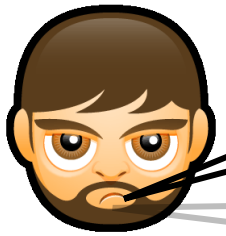
Round 8
-a -o -an, -ak -op -asp, -a -o -an, …

…

Experiment 2 participant

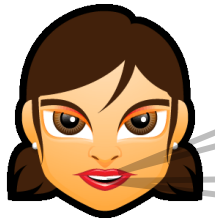


Experiment 1 participants



Round 2
-a -o -an, -a -op -asp, -a -o -an, …

Round 8
-a -o -an, -ak -op -asp, -a -o -an, …



Round 2
-ak -o -an, -ak -op -esp, -a -o -esp, …

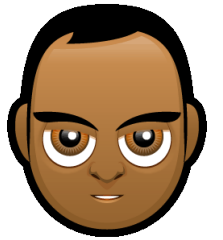
Round 8
-a -o -an, -ak -op -asp, -a -o -an, …

Round 2
-a -o -esp, -ak -o -esp, -a -o -esp, …

Round 8
-a -o -an, -ak -op -asp, -a -o -an, …

…

Experiment 2 participant



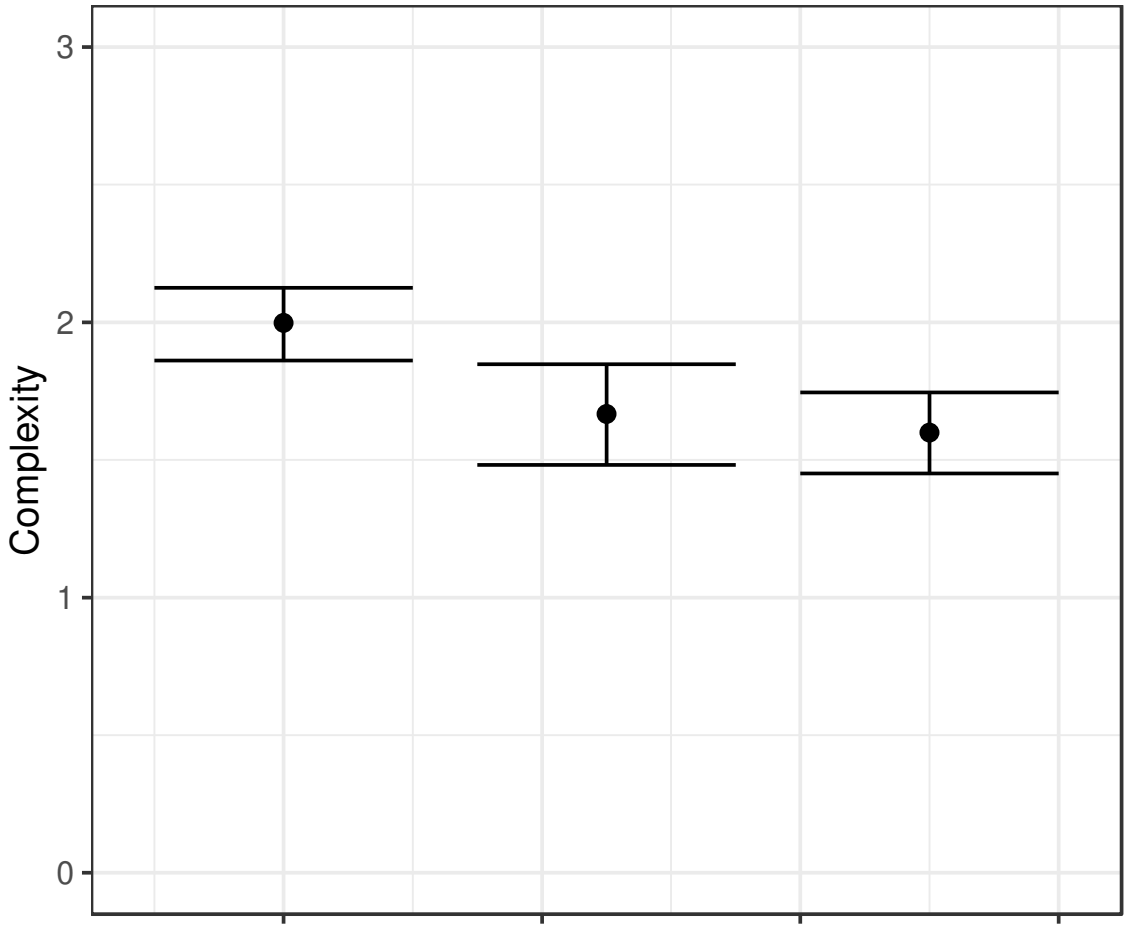
Experiment 2 (N=522)

Small (2) or Large (8) population

All Native, Mixed or All Non-native input

	All Native	**Mixed**	**All Non-native**
Small	Labels from 2 Round 8 learners	Labels from 1 Round 2 and 1 Round 8 learner	Labels from 2 Round 2 learners
Large	Labels from 8 Round 8 learners	Labels from 4 Round 2 and 4 Round 8 learners	Labels from 8 Round 2 learners

Experiment 2 results



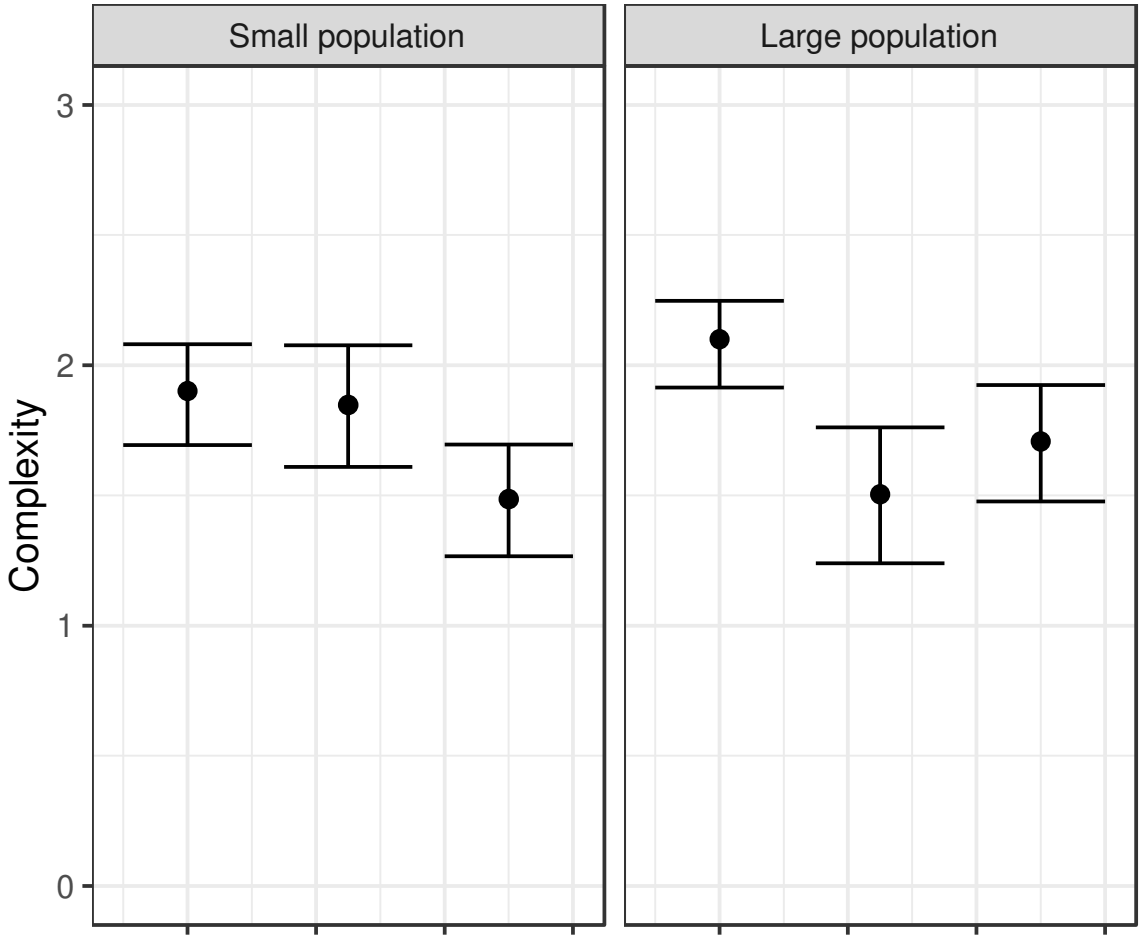
Data composition

- All Native
- Mixed
- All Non-native

Plot shows only round 8 data (N=222)
Stats use all N=522

All native > Mixed (b=-0.63, SE=0.22, p=.004)
Mixed ≈ All Non-native (b=-0.18, SE=0.22, p=.43
No effect of population size (b=0.16, SE=0.18, p=.38)
Interaction (b=0.96, SE=0.45, p=.031)

Experiment 2 results



Plot shows only round 8 data (N=222)
Stats use all N=522

Data composition

- All Native
- Mixed
- All Non-native

All native > Mixed (b=-0.63, SE=0.22, p=.004)
Mixed ≈ All Non-native (b=-0.18, SE=0.22, p=.43
No effect of population size (b=0.16, SE=0.18, p=.38)
Interaction (b=0.96, SE=0.45, p=.031)

Small population: Mixed-All Non-native contrast is significant
Large population: All Native-Mixed contrast is

Summary of Part 3

- Experiment 1: Learners simplify artificial language morphology (as you'd expect from parts 1-2), and early learners do so more
- Experiment 2: There are measurable consequences of learning from non-native-like early learners even in a single generation
- Learning from early learners / non-native speakers is a plausible mechanism driving the correlation between population size, proportion of non-native speakers, and language complexity
- In relation to part 1: learning pressure depends on population characteristics and might therefore differ in different populations

Conclusions

- Language is transmitted via repeated **learning** and **use**
- Language is shaped by these processes
- These processes create the regular compositional structure that makes language so powerful
- We can understand the distribution of **exceptions** to those regularities (within and across languages) in terms of those same processes