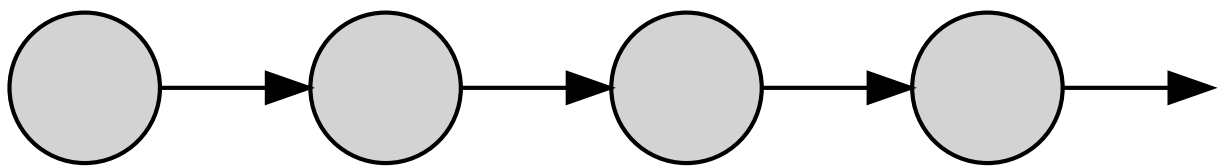
Cultural evolution

Using evolutionary principles to make sense of
word origins,
problem solving, and
the growth of Wikipedia articles.

Pierce Edmiston
pedmiston@wisc.edu

Unifying theme

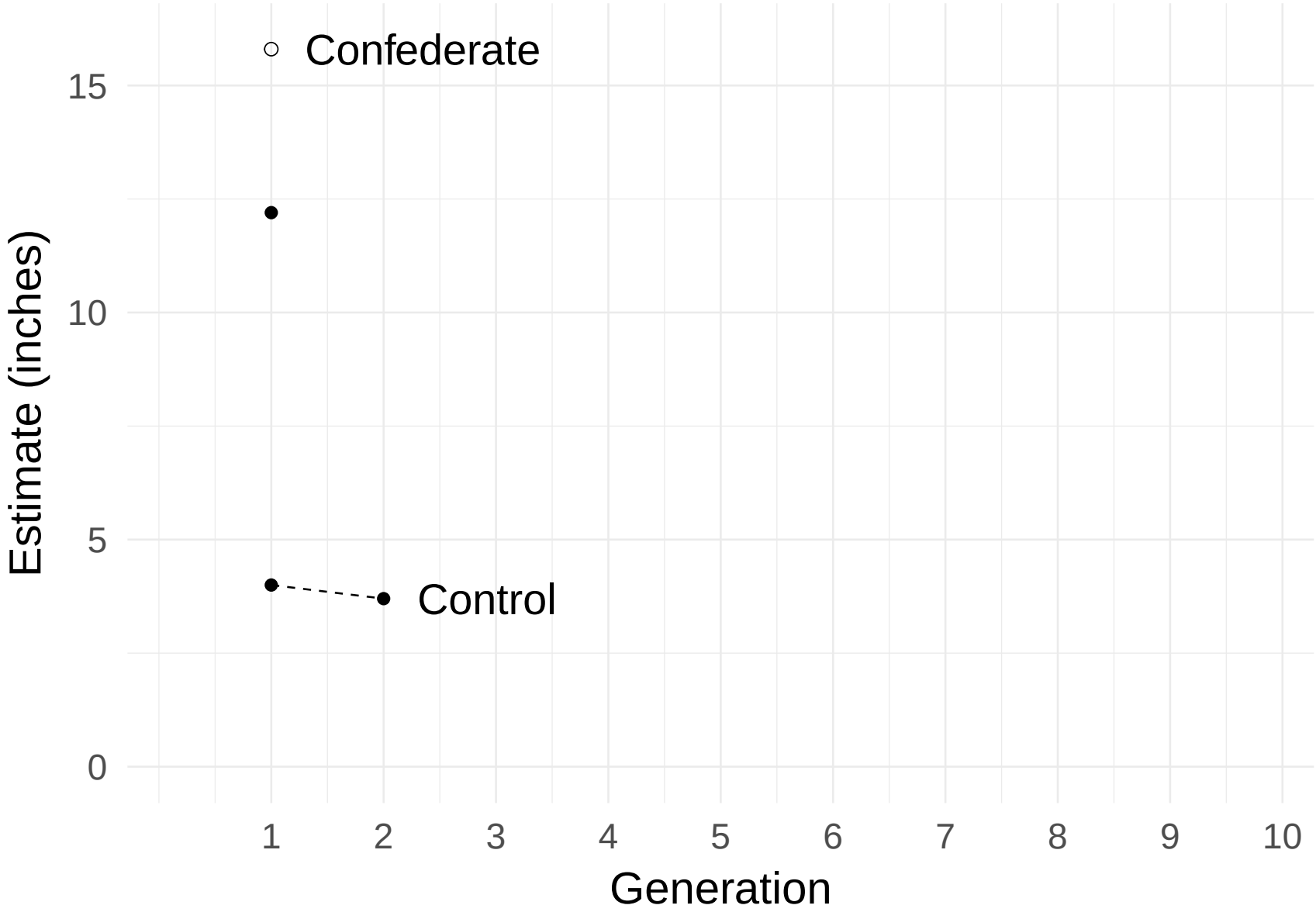
- **Iteration** is “the repetition of a process or utterance.”
- When can iteration be trusted?



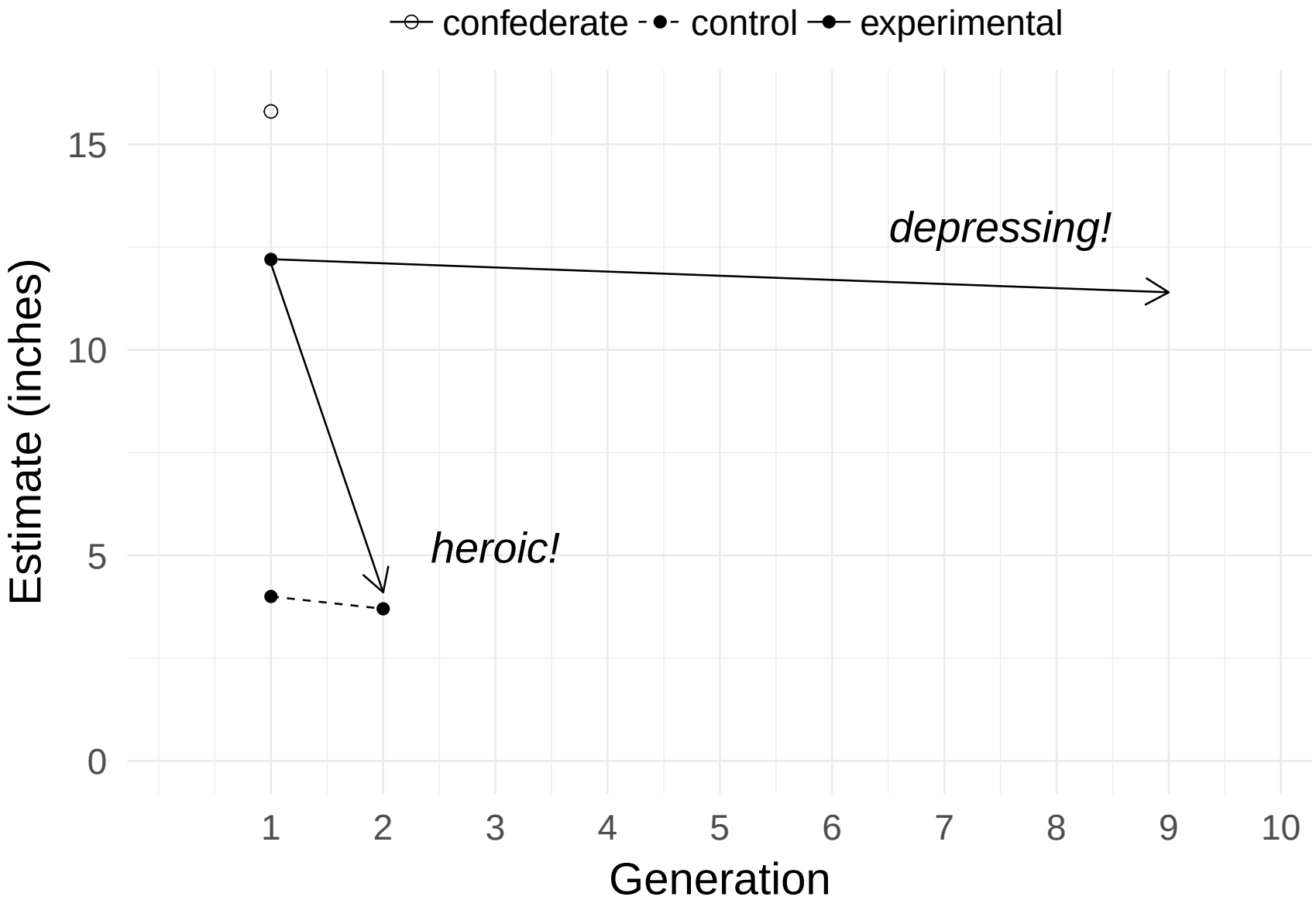
Evolution in the psychology department

Jacobs & Campbell. (1961). *J Abnorm Soc Psychol*.
The perpetuation of an arbitrary tradition through several generations of a laboratory microculture.

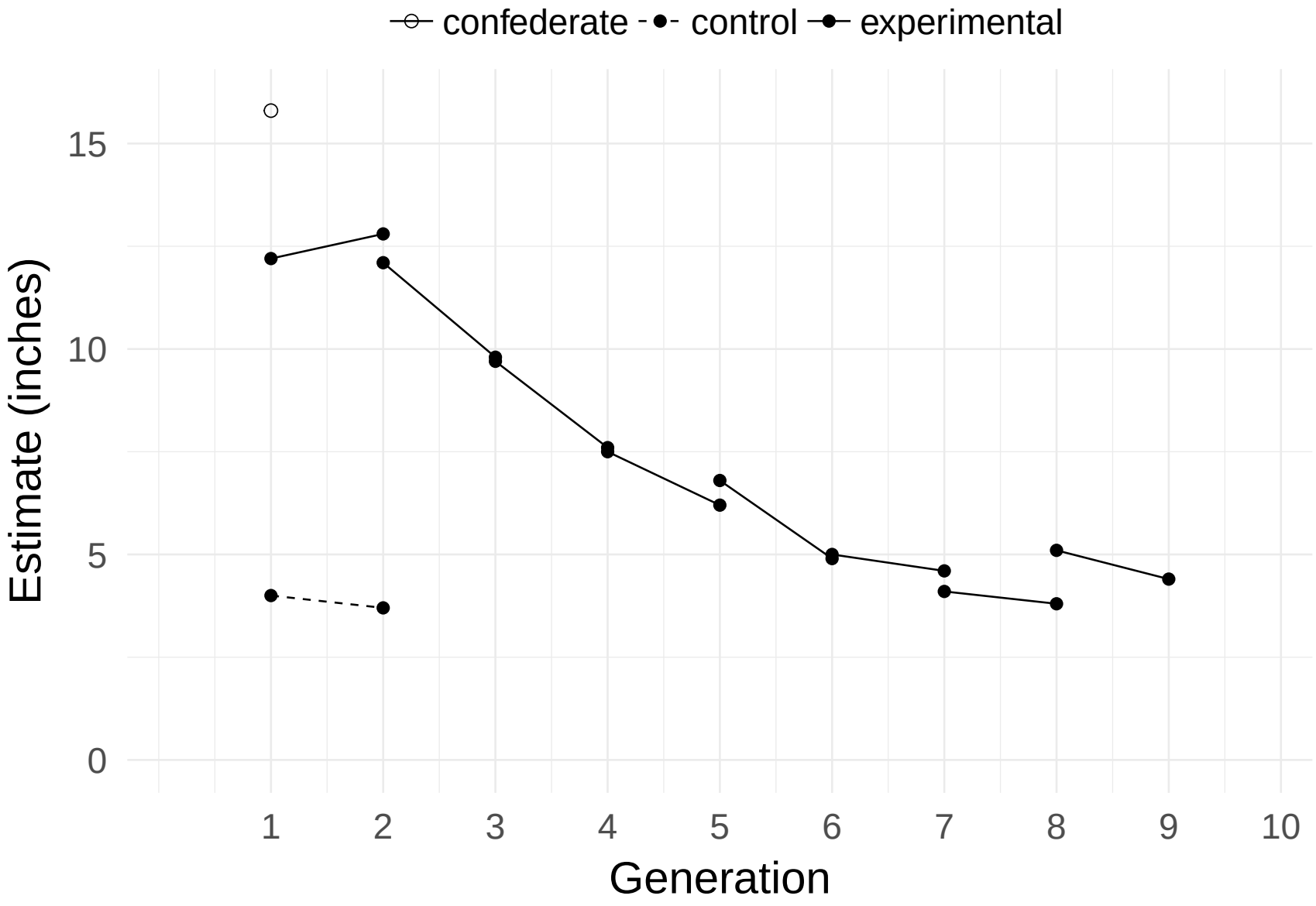
Iterated conformity (Jacobs & Campbell, 1961)



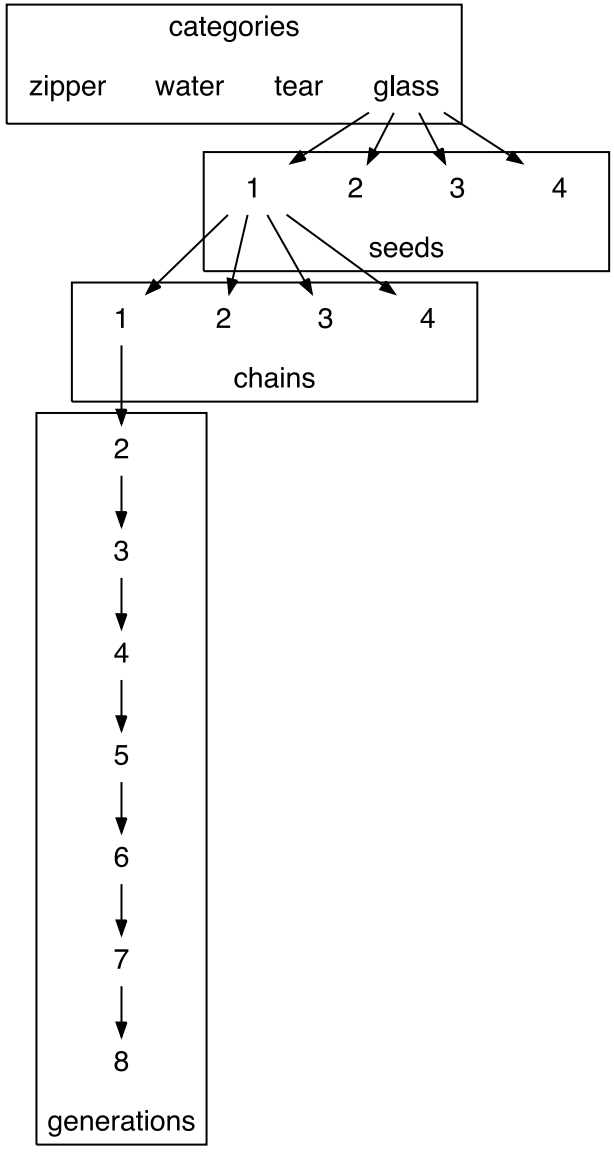
Iterated conformity (Jacobs & Campbell, 1961)



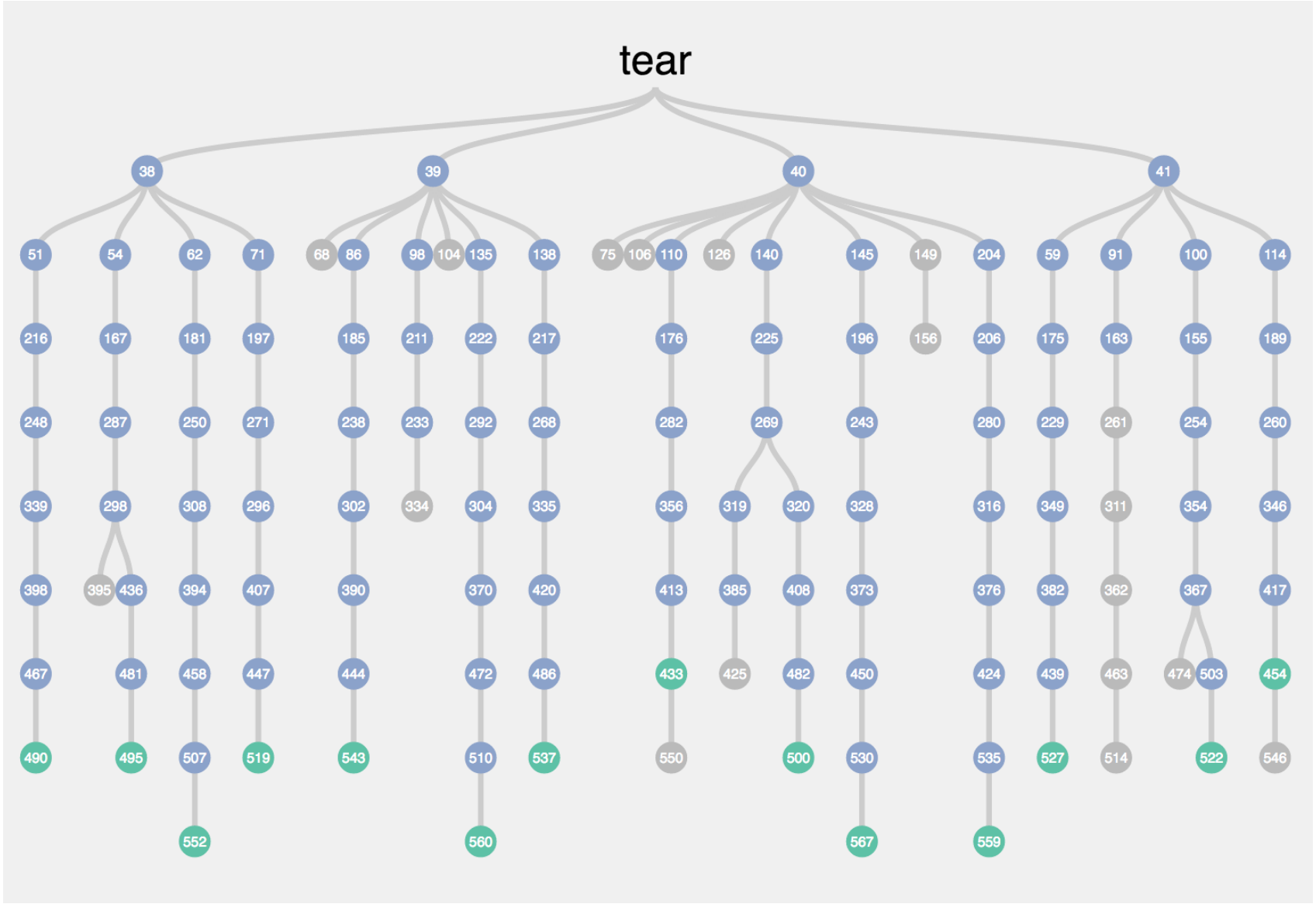
Iterated conformity (Jacobs & Campbell, 1961)



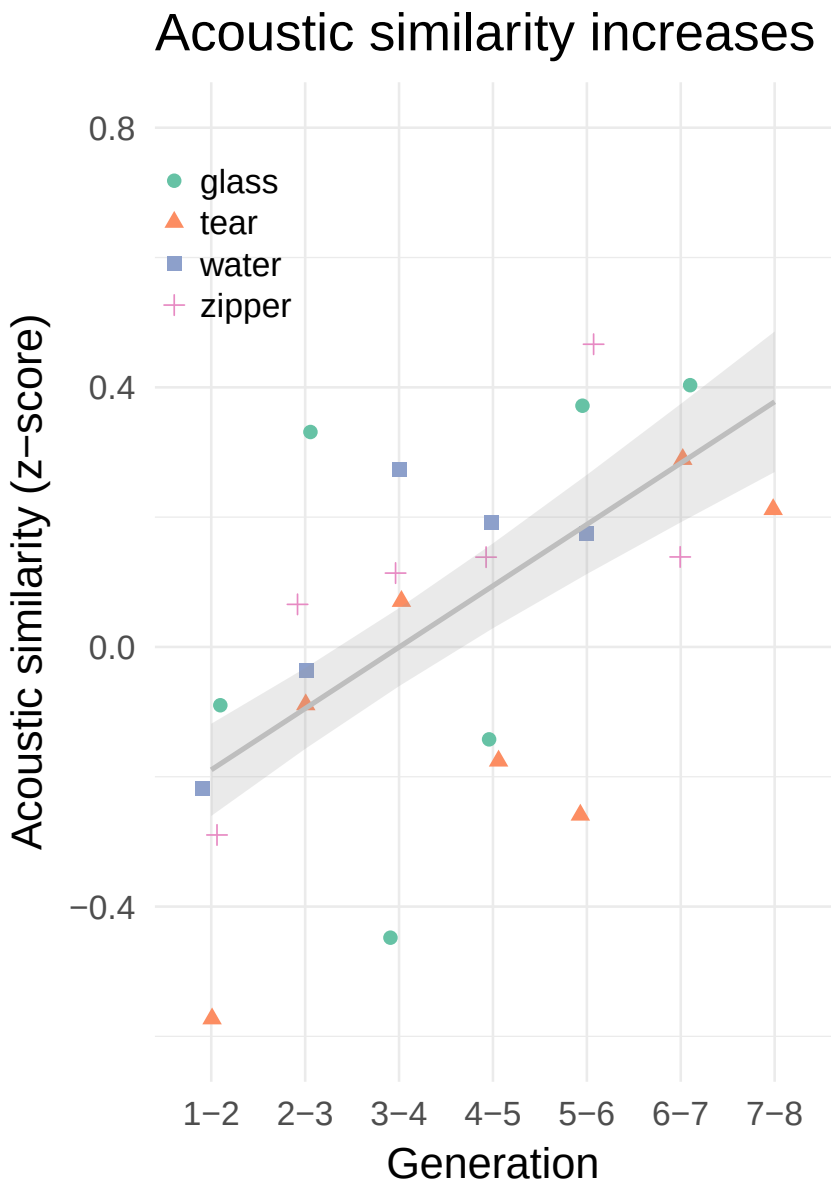
Creating words from iterated vocal imitation

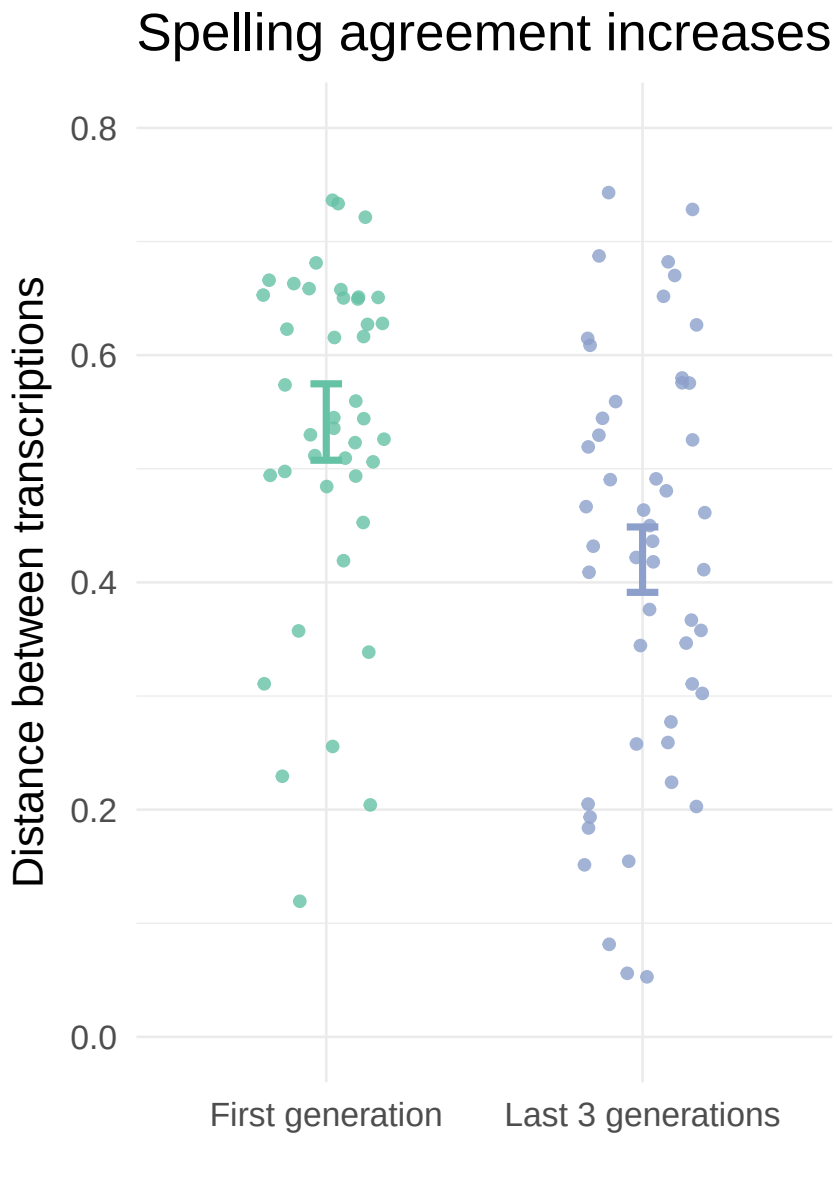


Telephone app

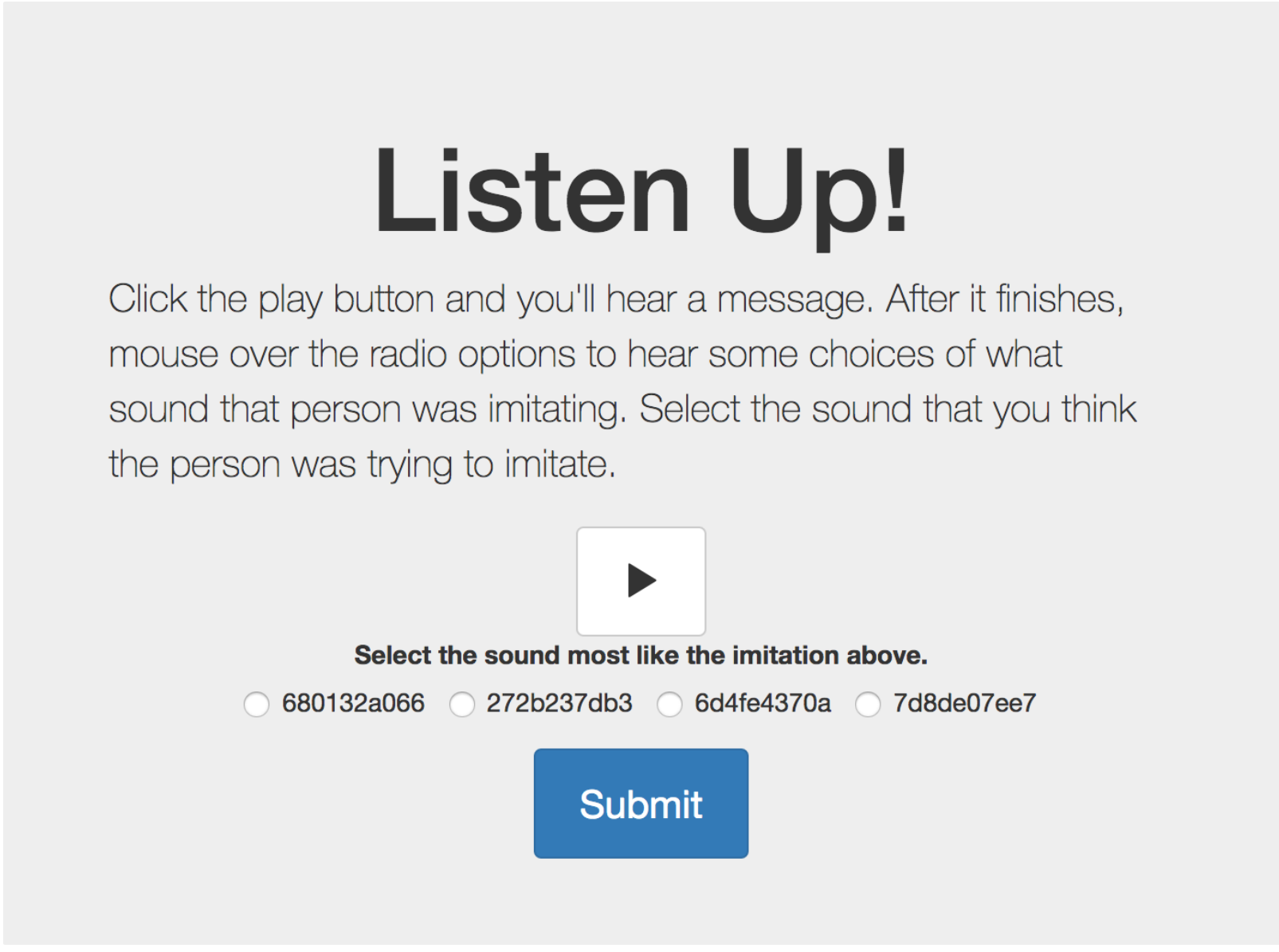


Imitations stabilize over generations

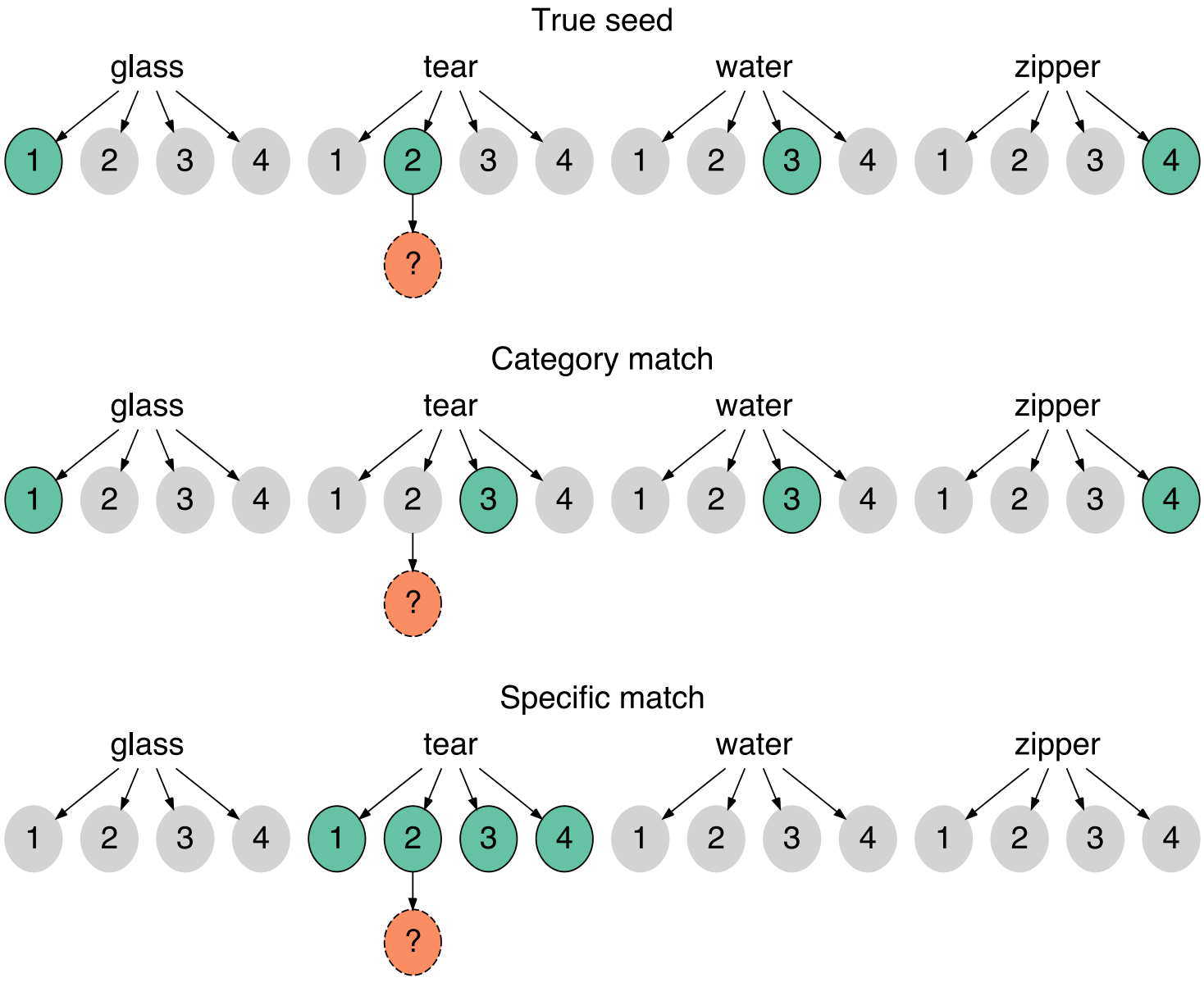




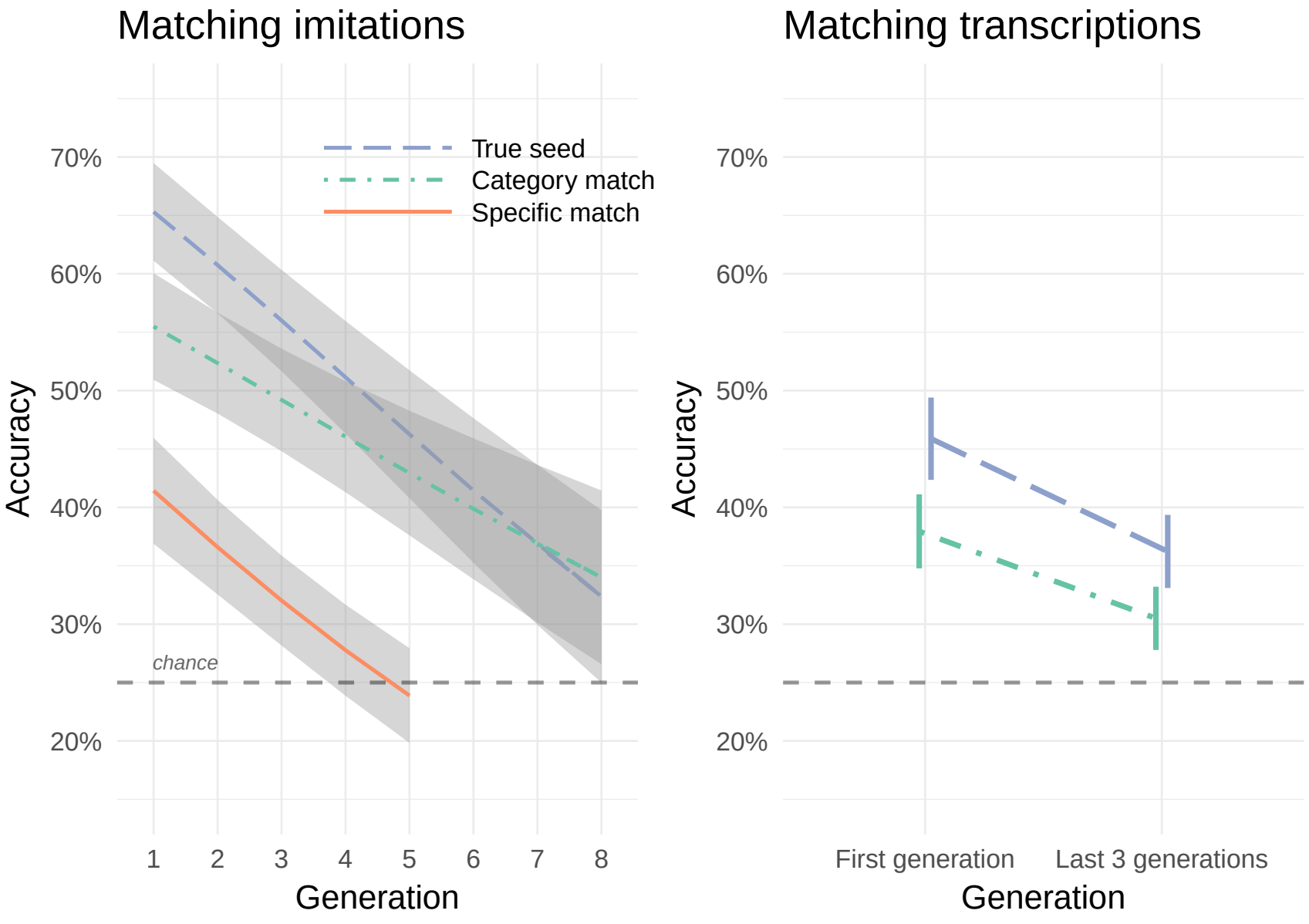
“Guess the seed” game



“Guess the seed” question types



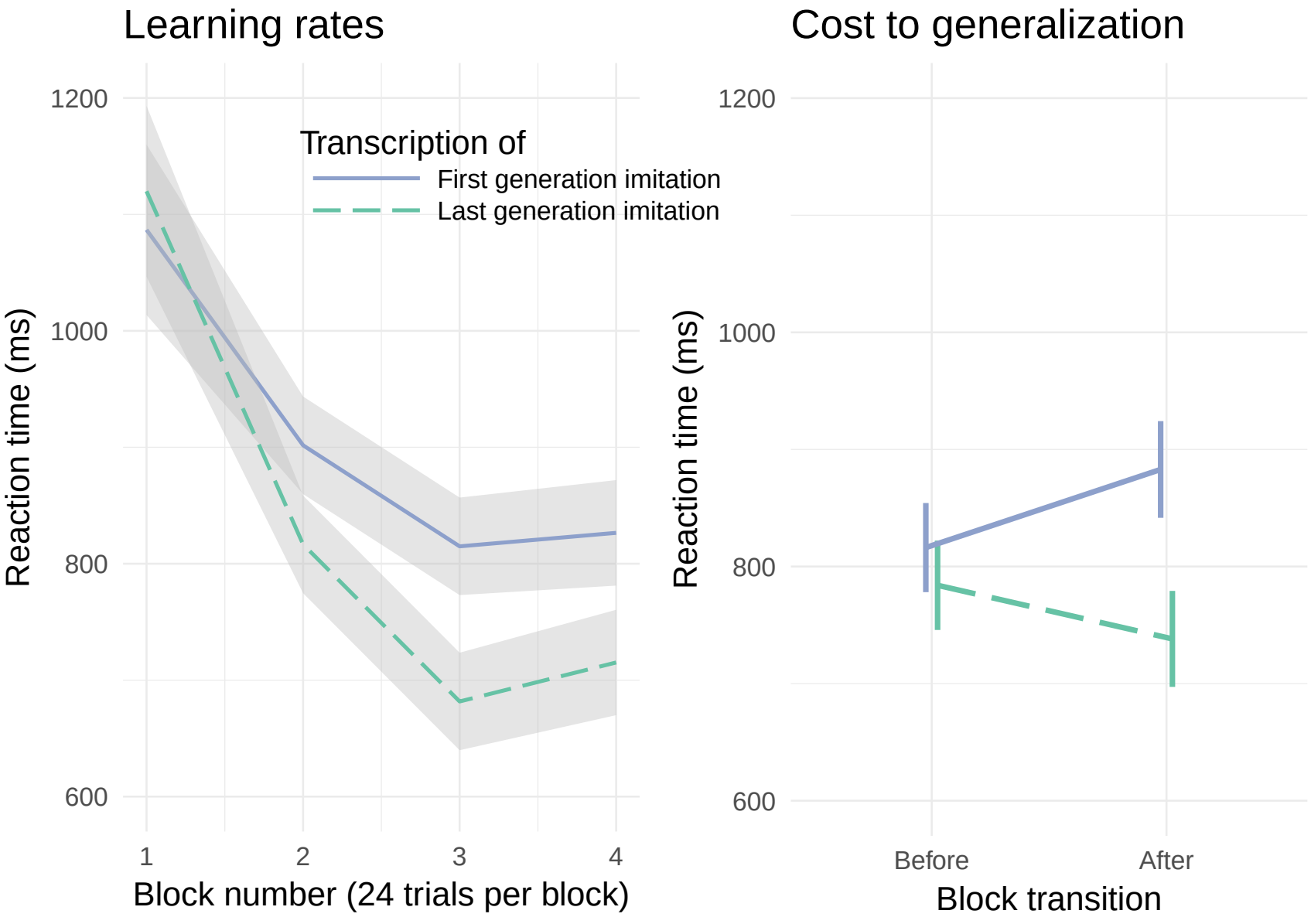
Matching accuracy



Invented words

Category	Seed	First generation	Last generation
glass	1	tingtingting	dundunduh
glass	2	chirck	correcto
glass	3	dirrng	wayew
glass	4	boonk	baroke
tear	1	scheeept	cheecheea
tear	2	feeshefee	cheeoooo
tear	3	hhhweerrr	chhhhhhewwwe
tear	4	ccccchhhhyeaahh	shhhhh
water	1	boococucuwich	eeverlusha
water	2	chwoochwooochwooo	cheiopshpshcheiopsh
water	3	atoadelchoo	mowah
water	4	awakawush	galonggalong
zipper	1	euah	izoo
zipper	2	zoop	veeeep
zipper	3	arrgt	owww
zipper	4	bzzzzup	izzip

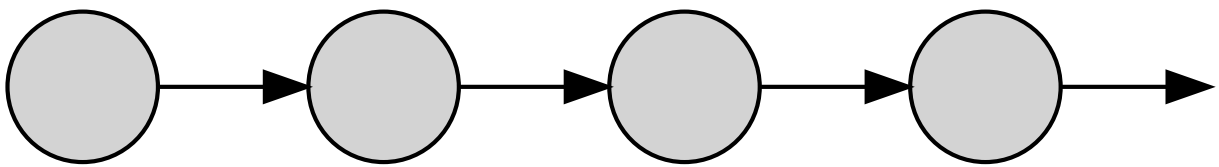
Category learning



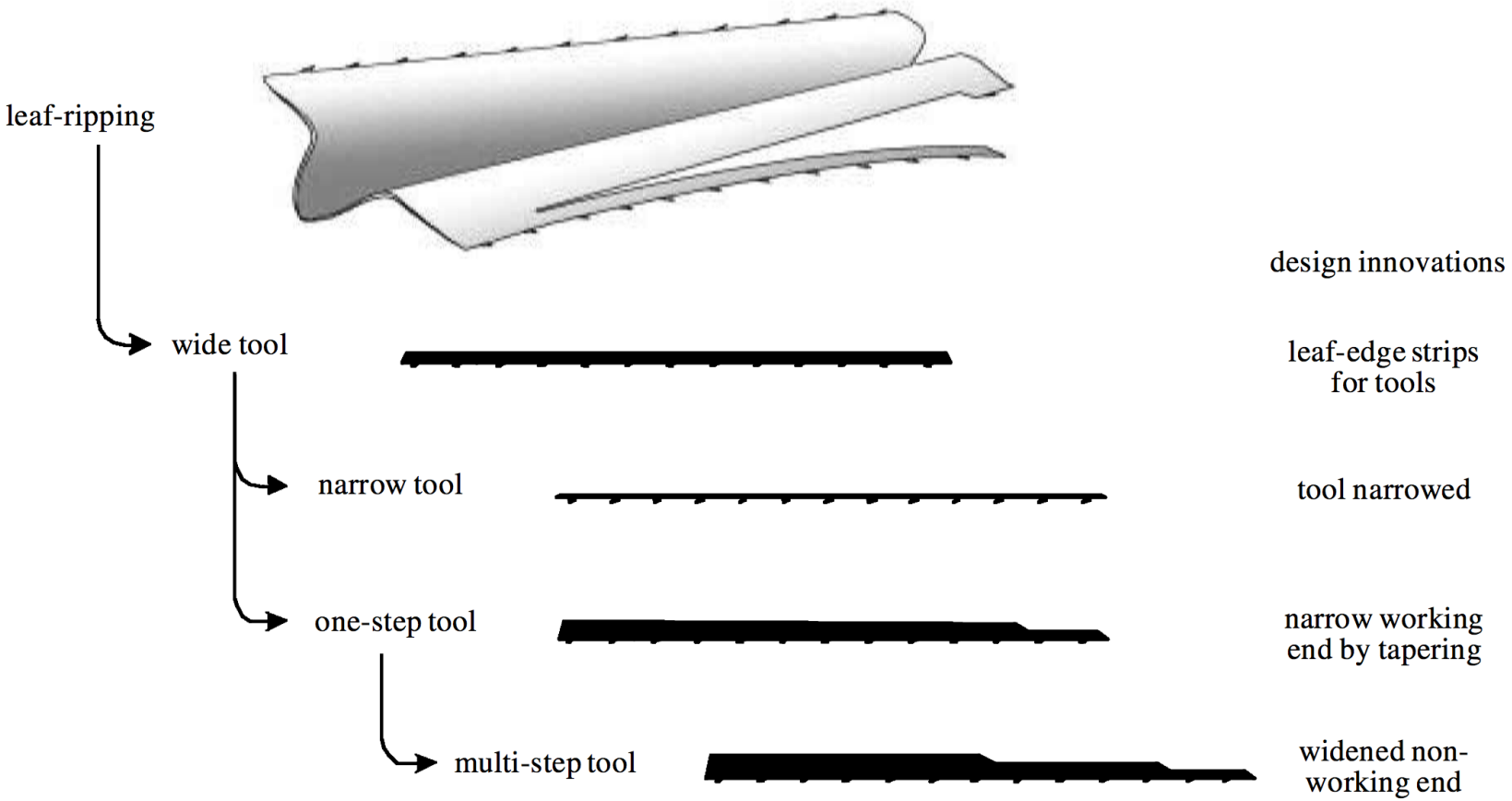
Summary: Word origins

Unguided repetition of nonverbal imitations makes them more word-like.

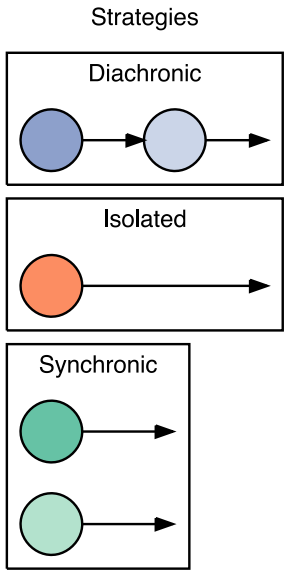
- Acoustic form becomes more repeatable and easier to spell.
- Imitations and transcriptions gradually lose resemblance to source.
- As imitations transition into words they become more categorical.

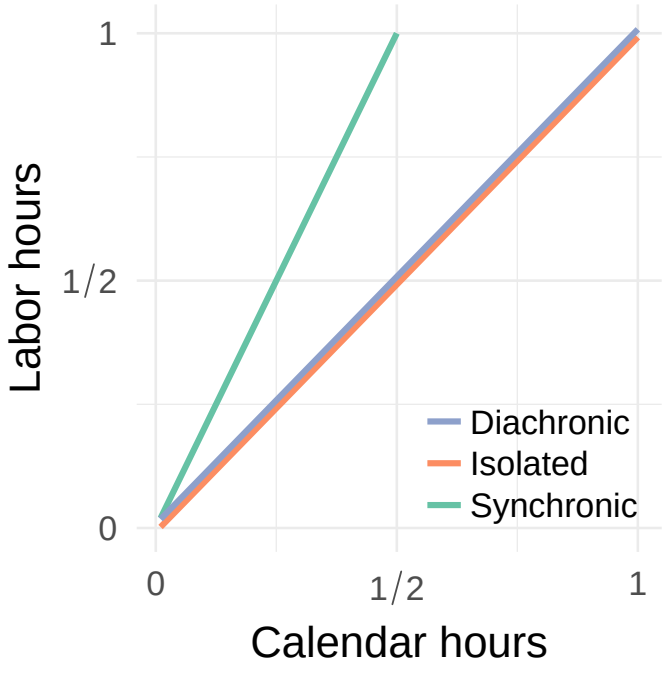


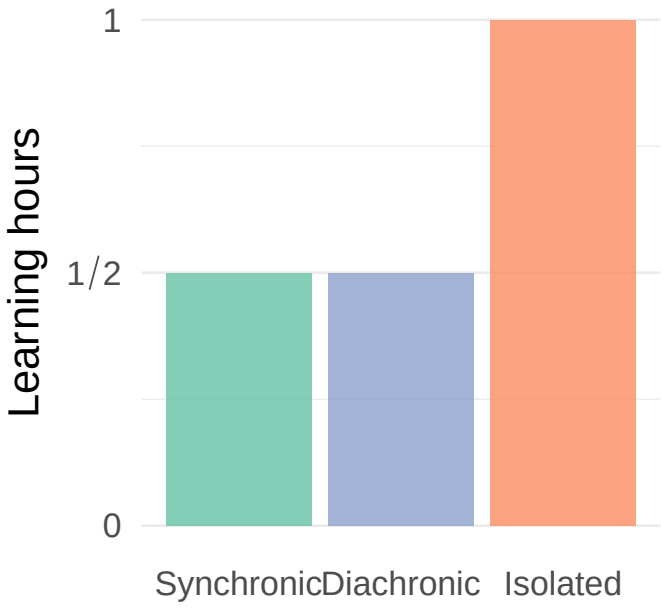
Technological evolution (Hunt & Gray, 2003)



Strategies

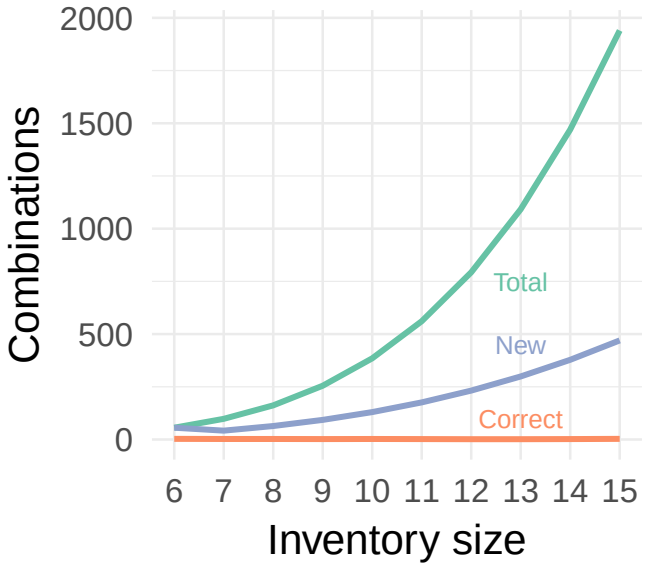

Strategies
Diachronic
Isolated
Synchronic

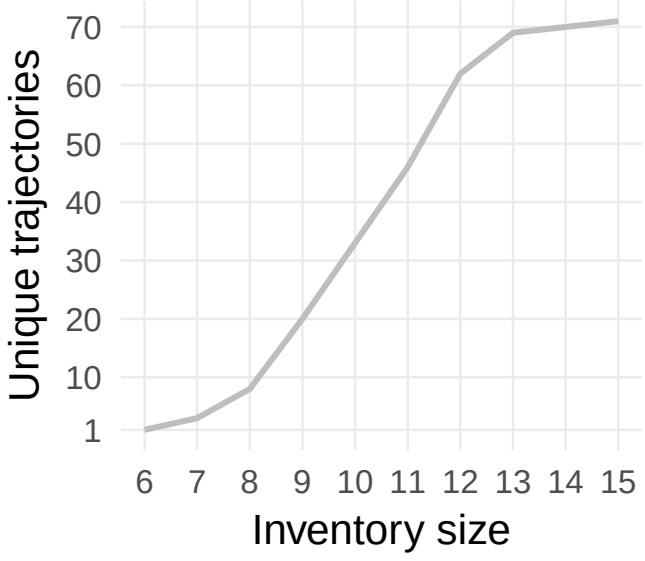

1
1/2
0
Labor hours
0
1/2
1
Calendar hours
Diachronic
Isolated
Synchronic


1
1/2
0
Learning hours
Synchronic
Diachronic
Isolated

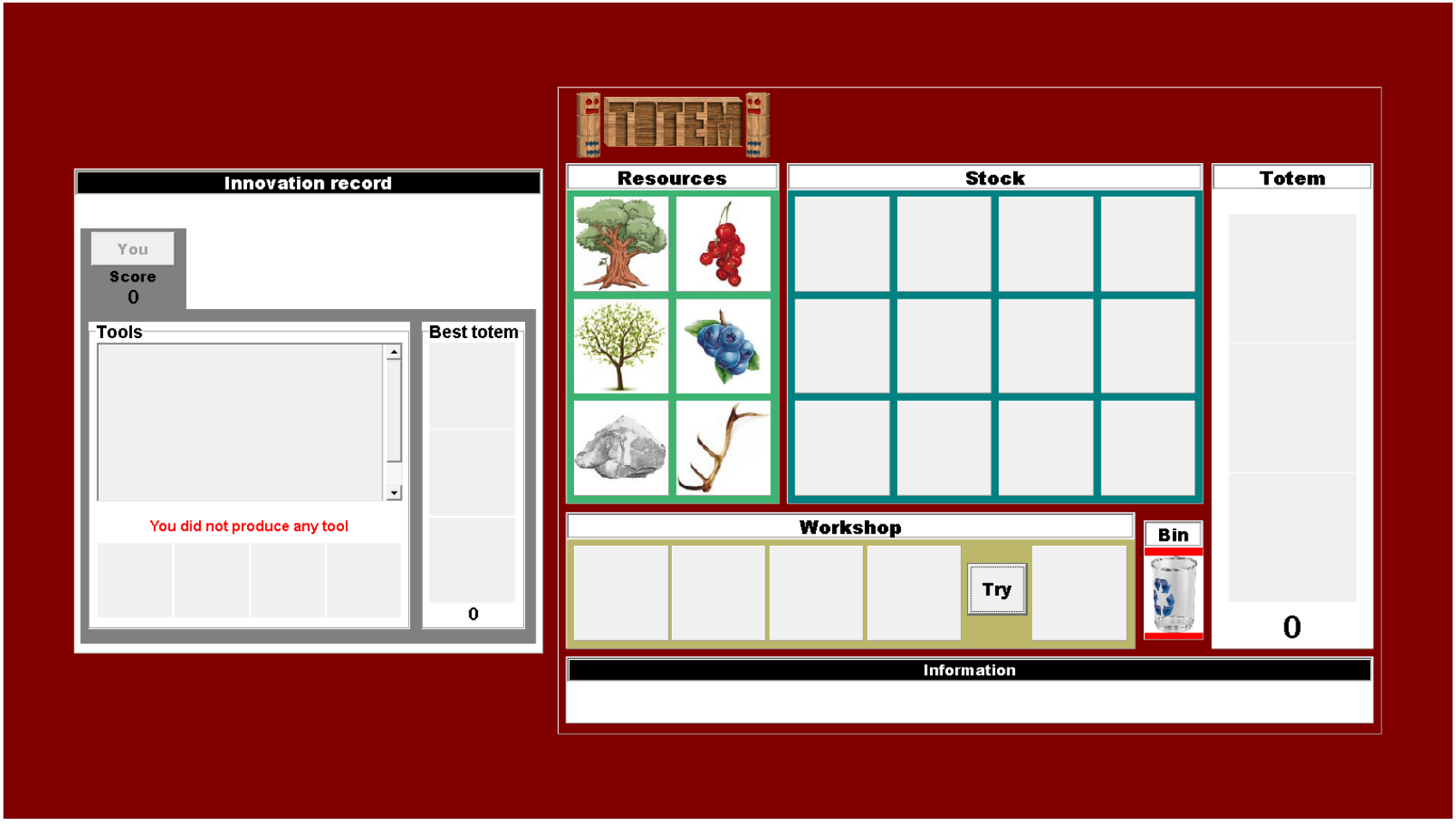
Innovation problem solving (Derex & Boyd, 2015)

Innovation landscape

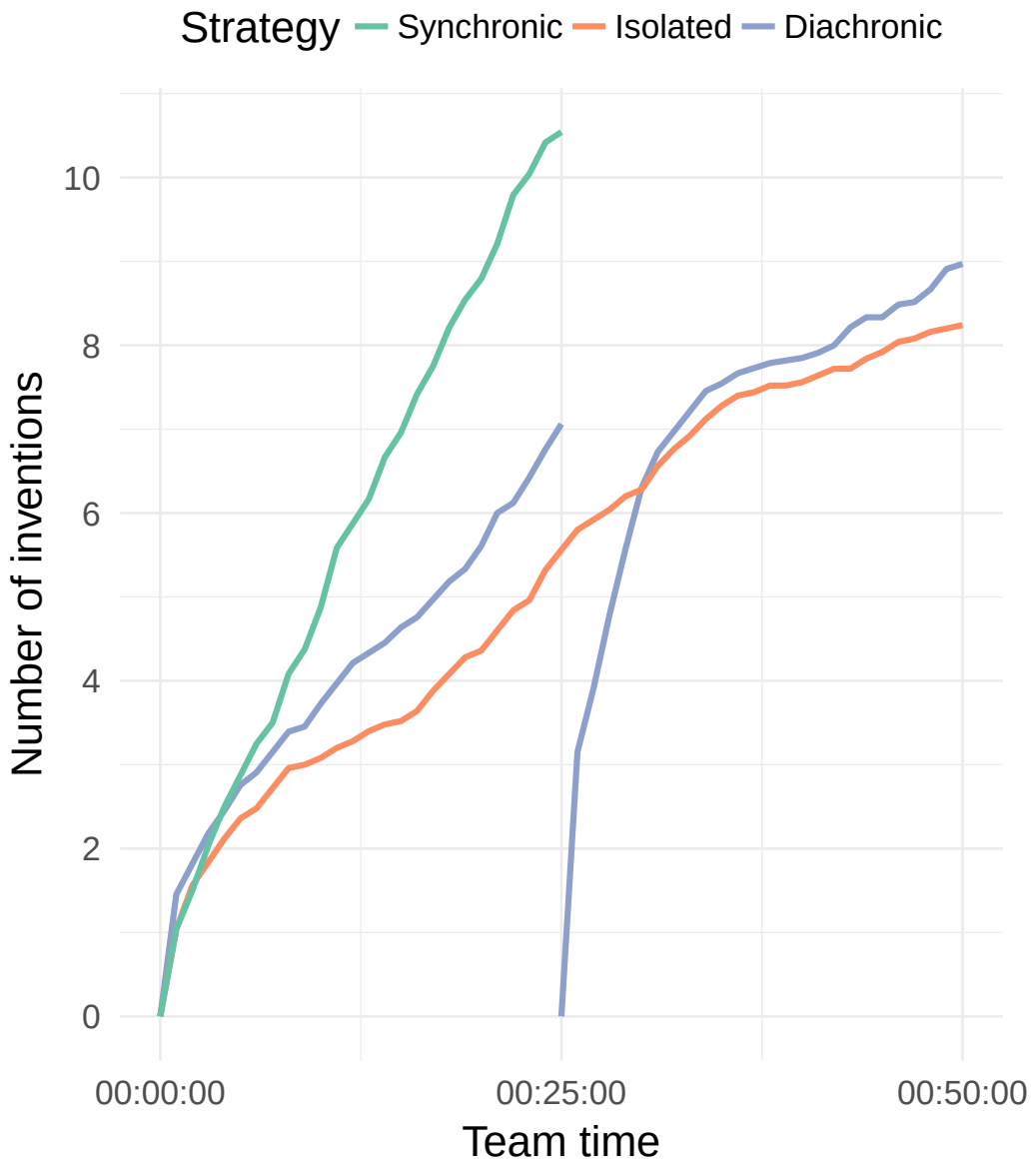


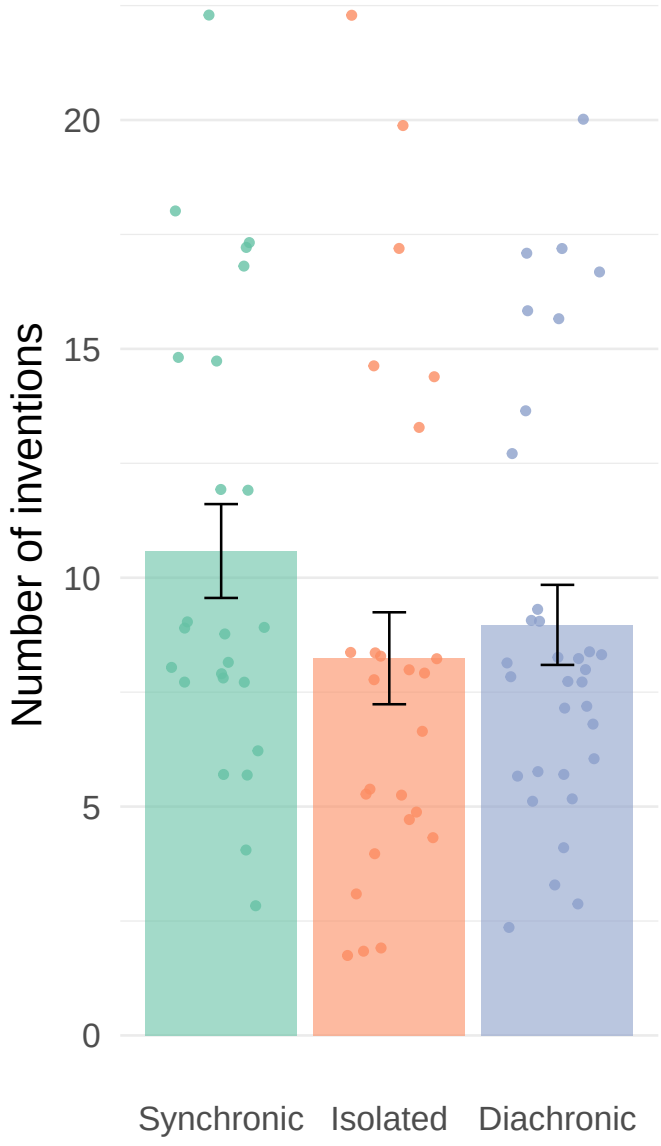


Totems game


Innovation record
You
Score
0
Tools
You did not produce any tool
Best totem
0
TOTEM
Resources
Stock
Totem
Workshop
Try
Bin
0
Information

Number of inventions

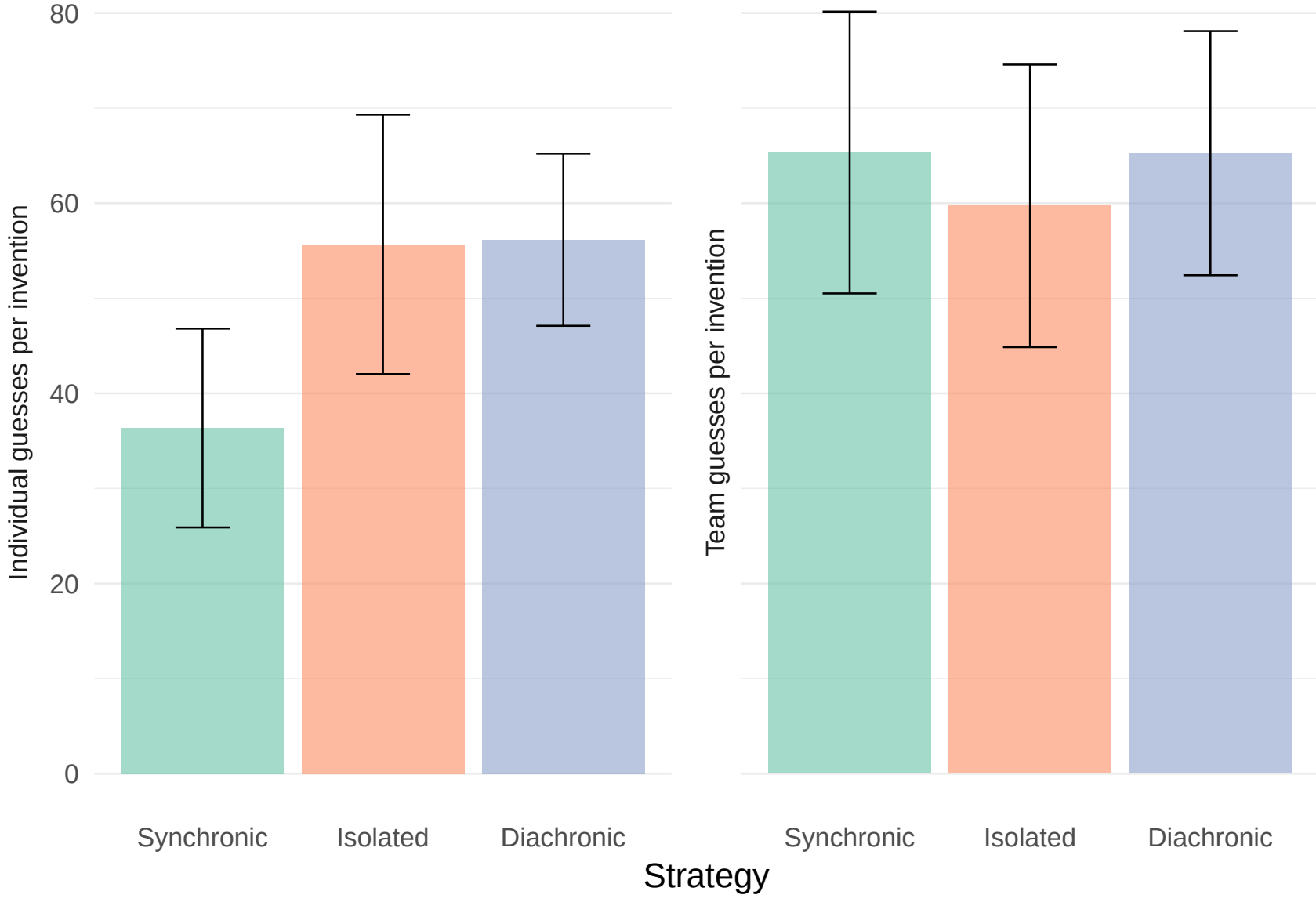




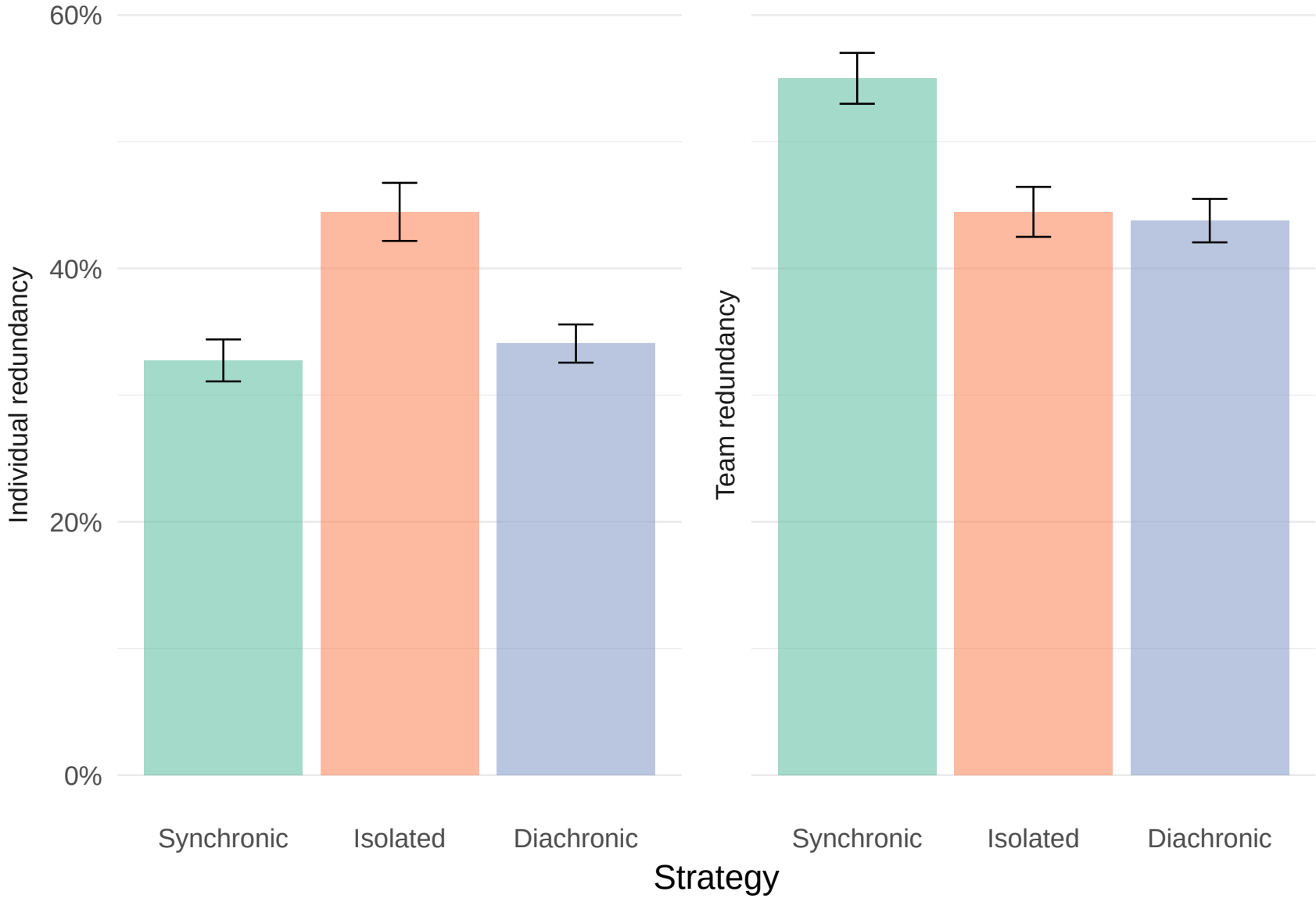
Measuring differences in problem solving

- Effectiveness (guesses per invention)
- Redundancy (non-unique guesses)
- Trajectories (unique paths)

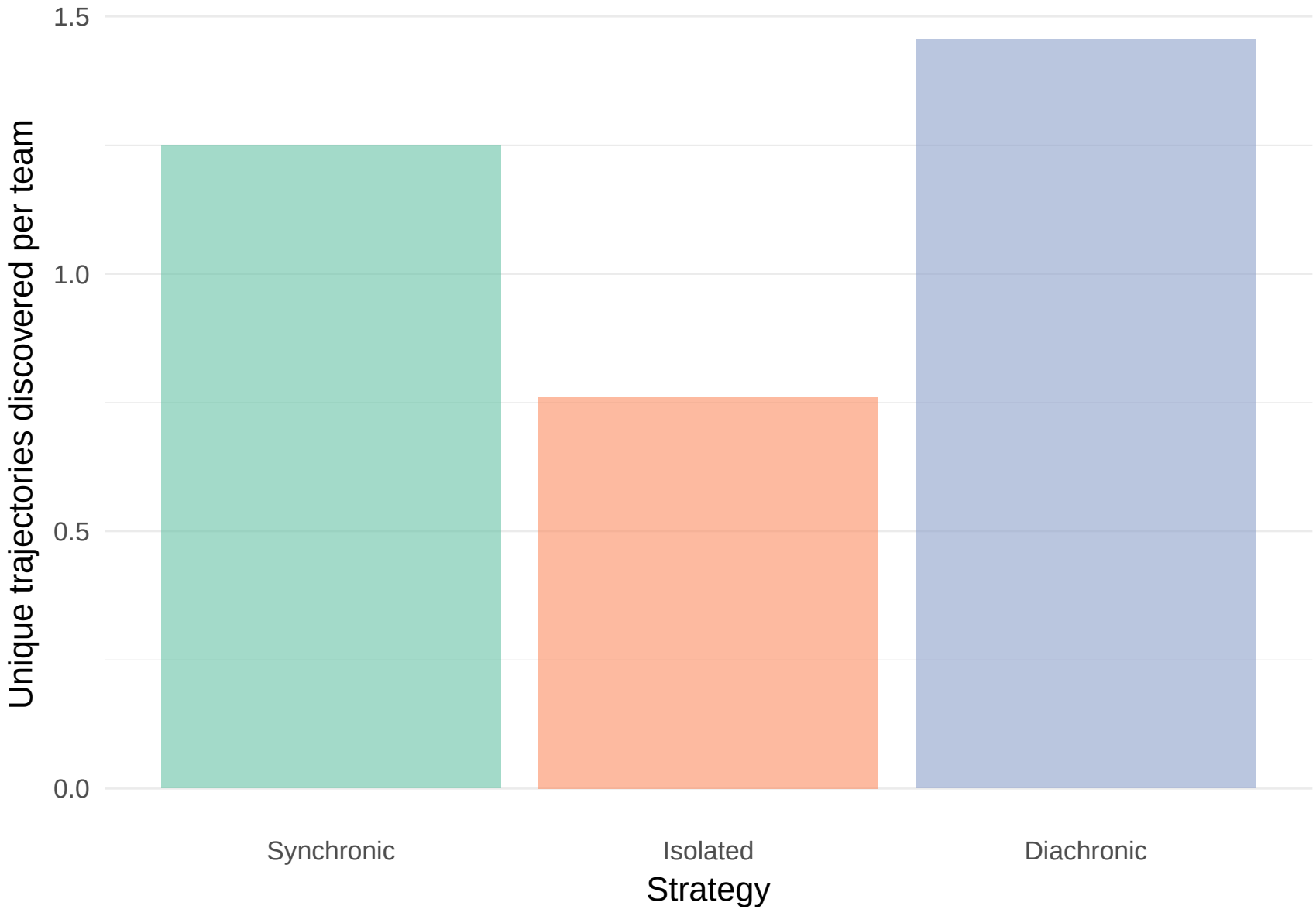
Effectiveness: Guesses per invention



Redundancy: Non-unique guesses

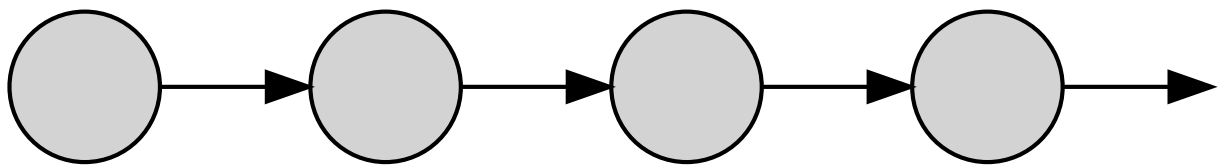


Trajectories: Exploration of landscape

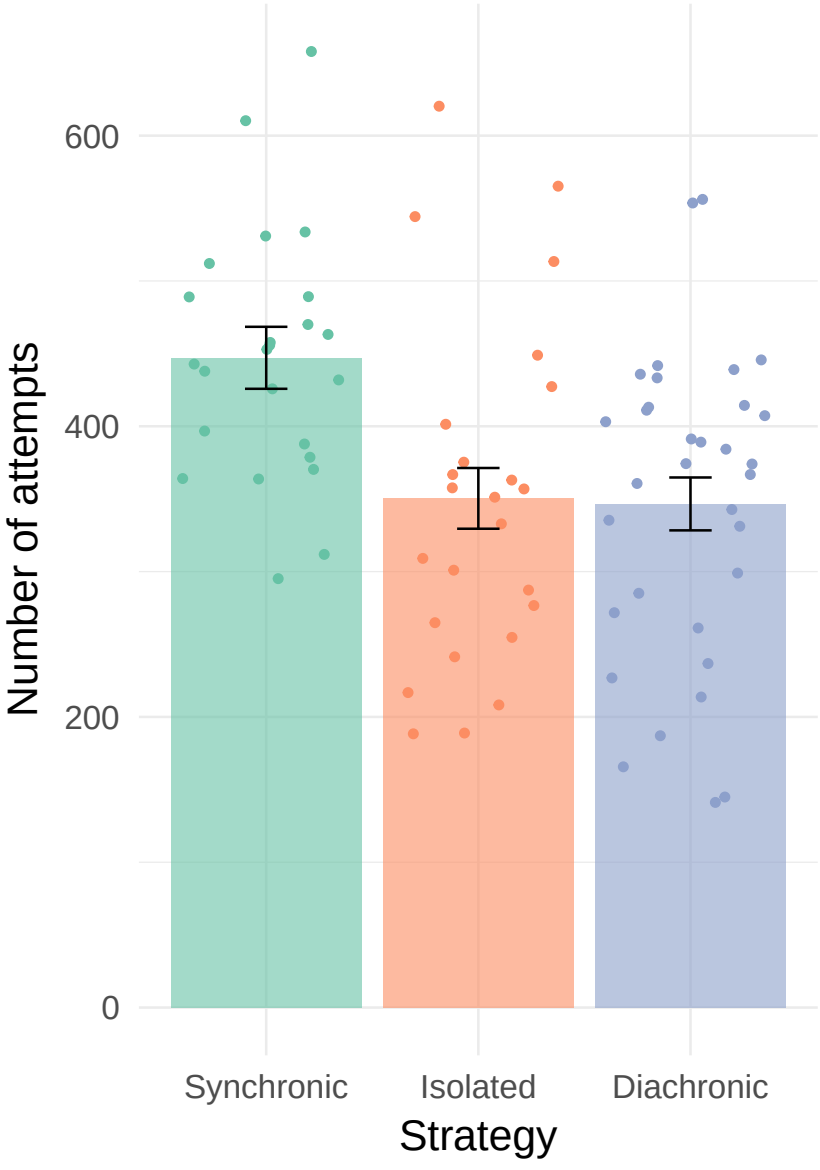


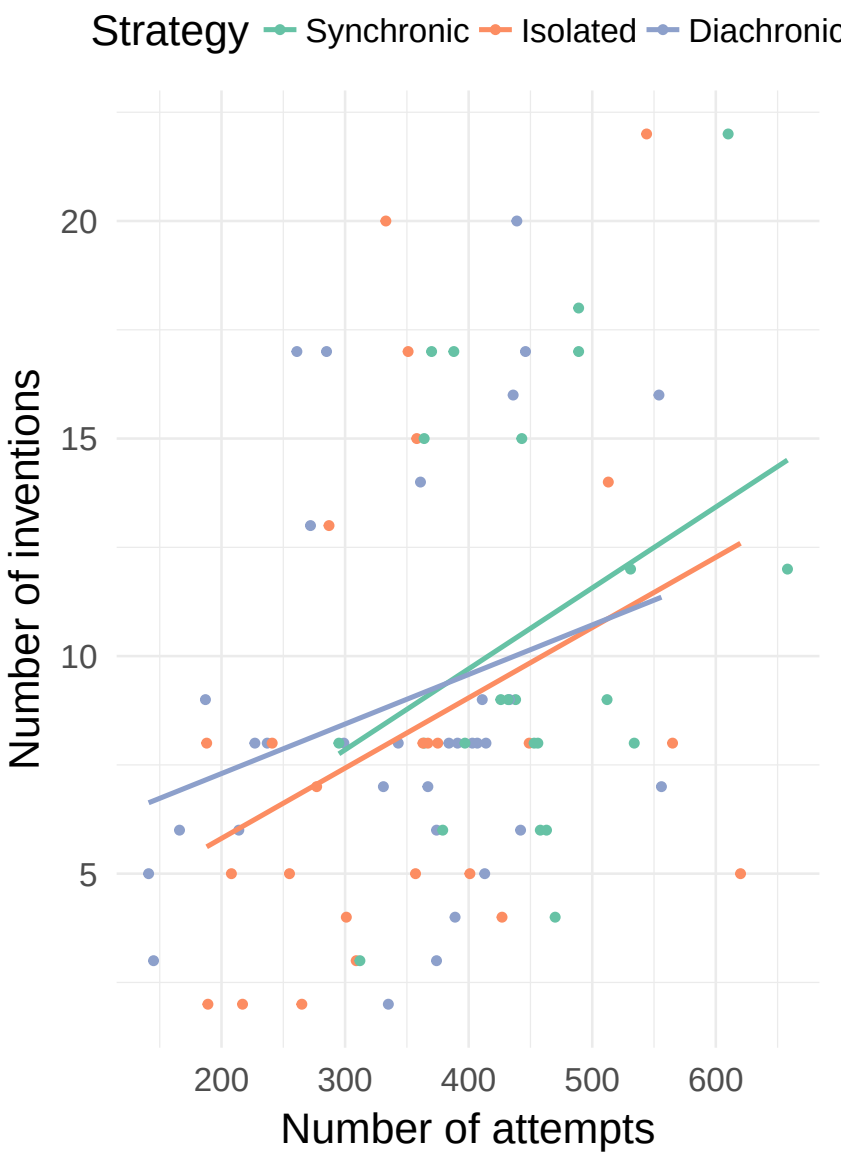
Summary: Diachronic inheritance

- Diachronic teamwork is the least redundant.
- Synchronic teamwork results in the most attempts.

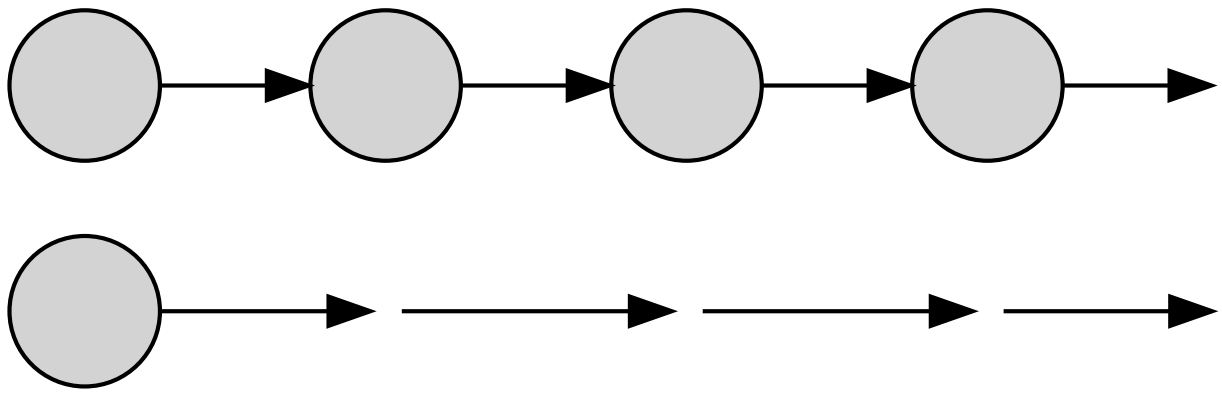


Total attempts

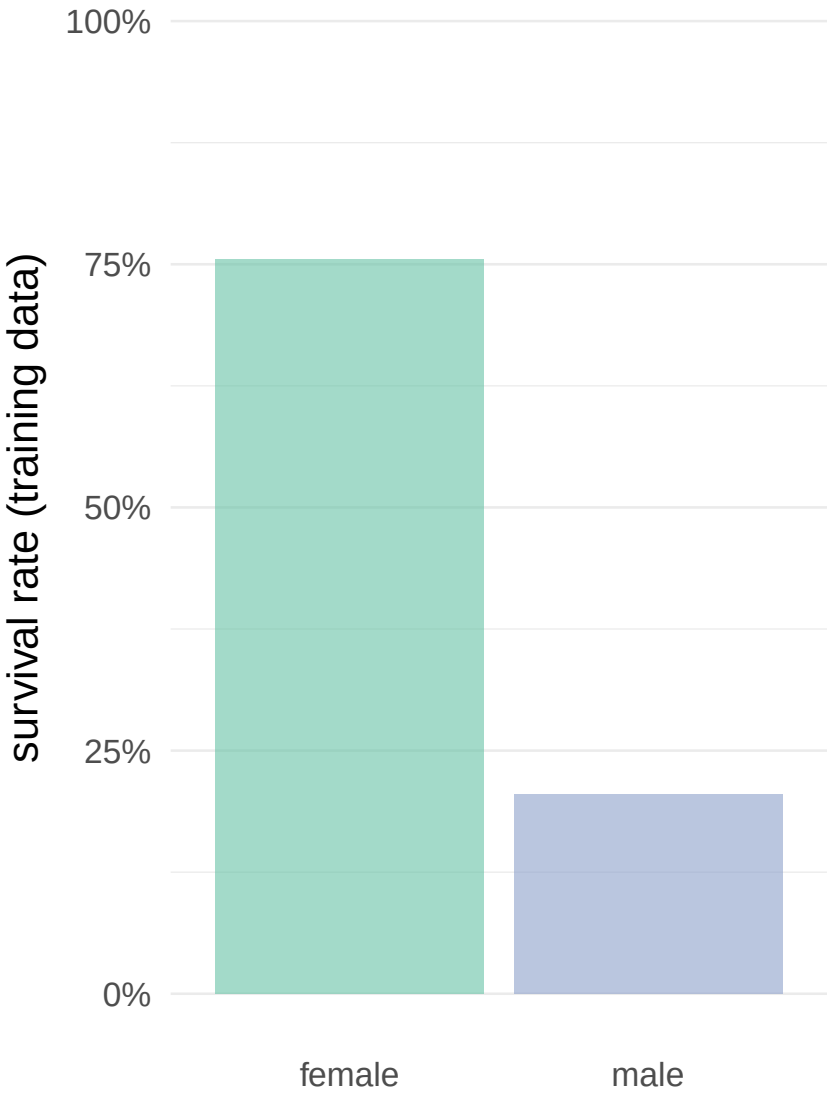


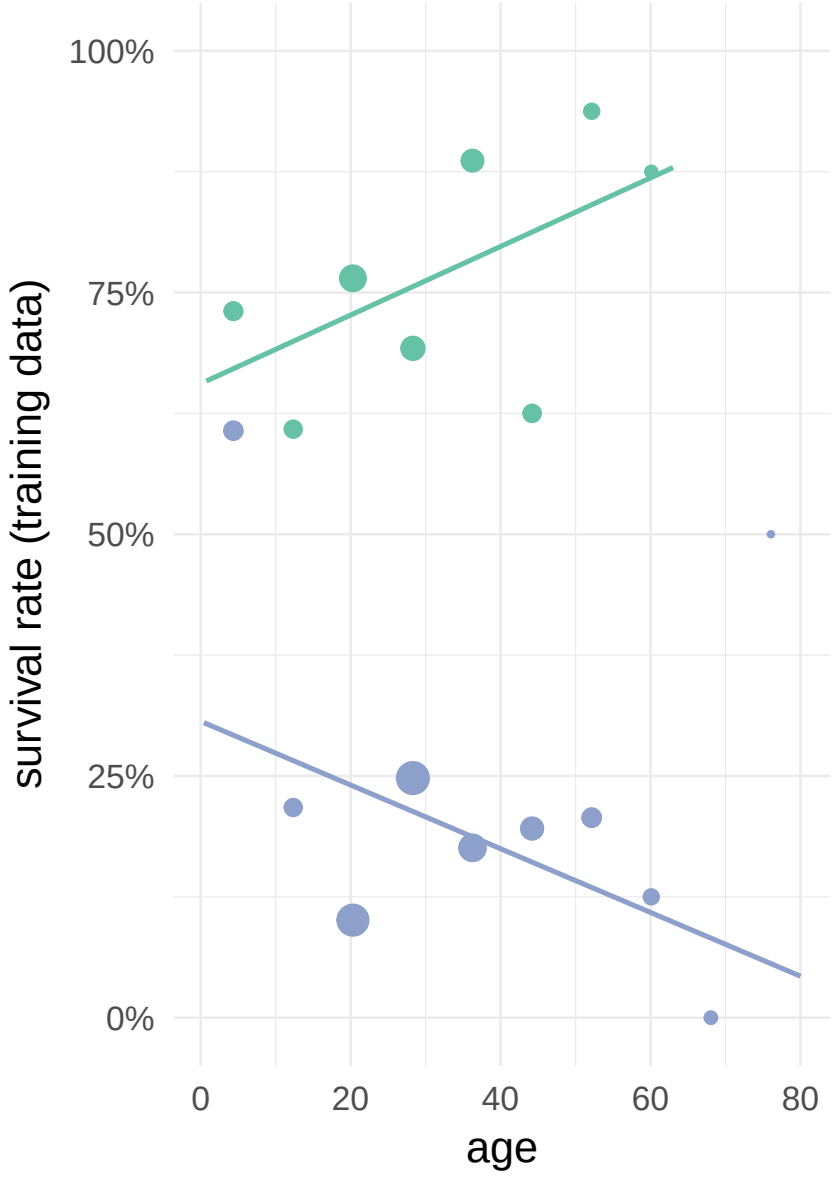


Iteration versus inheritance

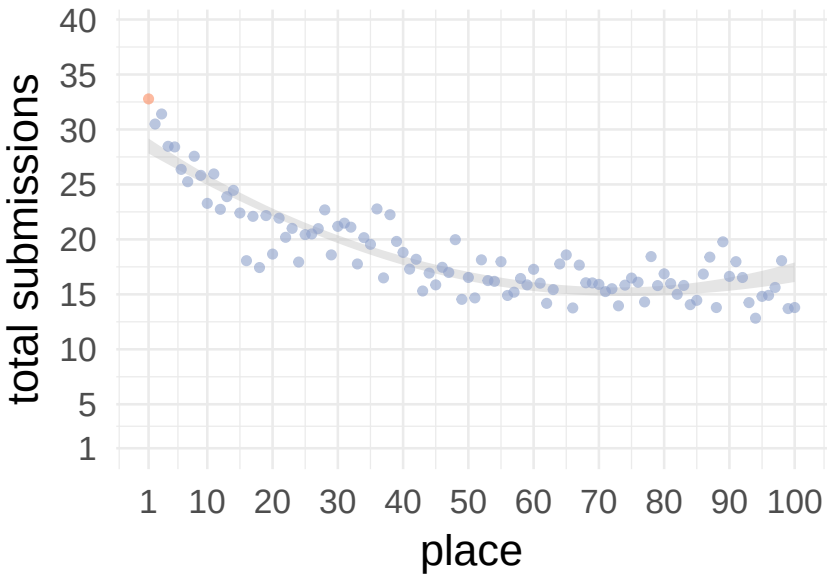


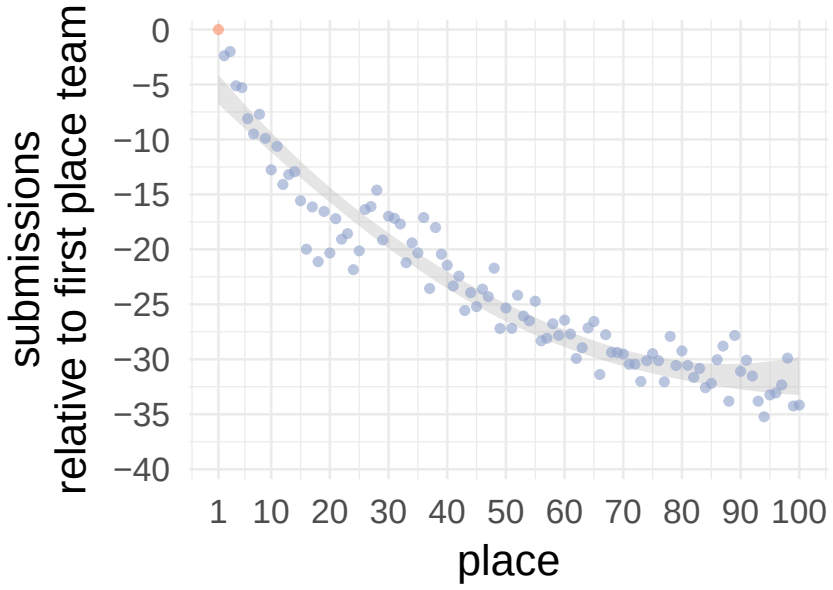
Kaggle competitions

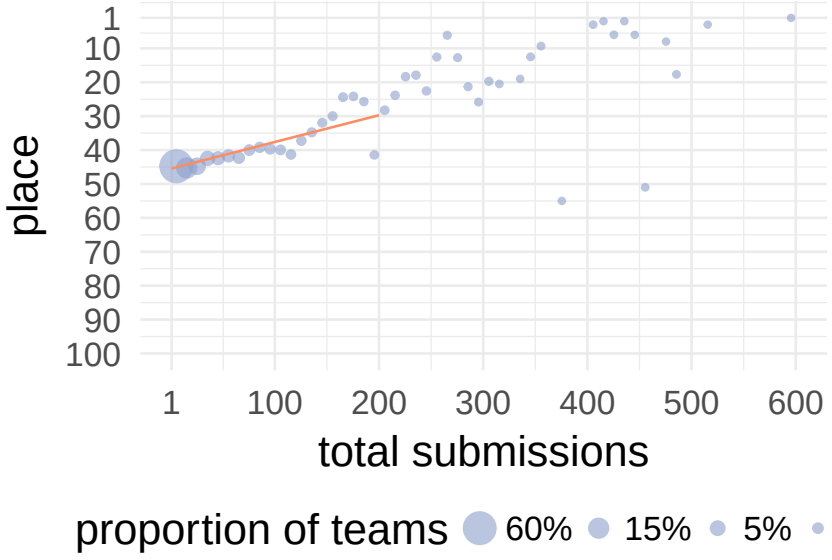


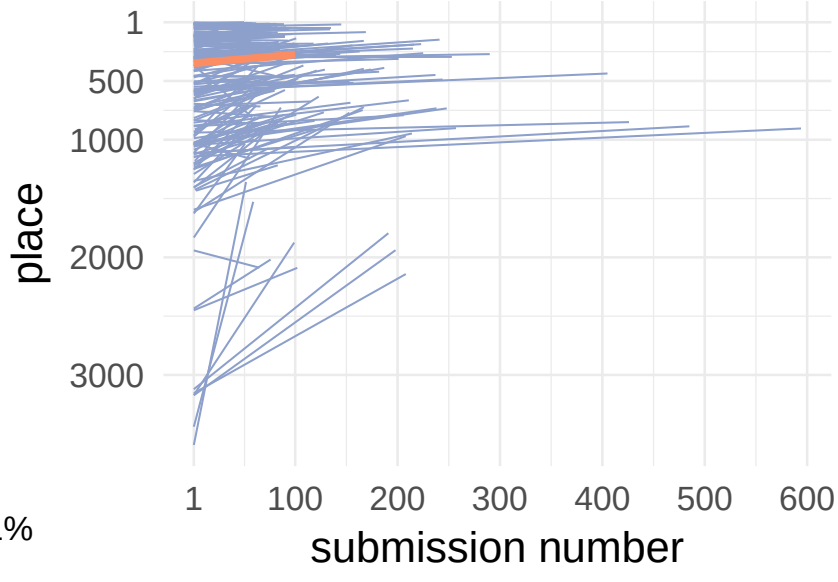


Kaggle competition leaderboards

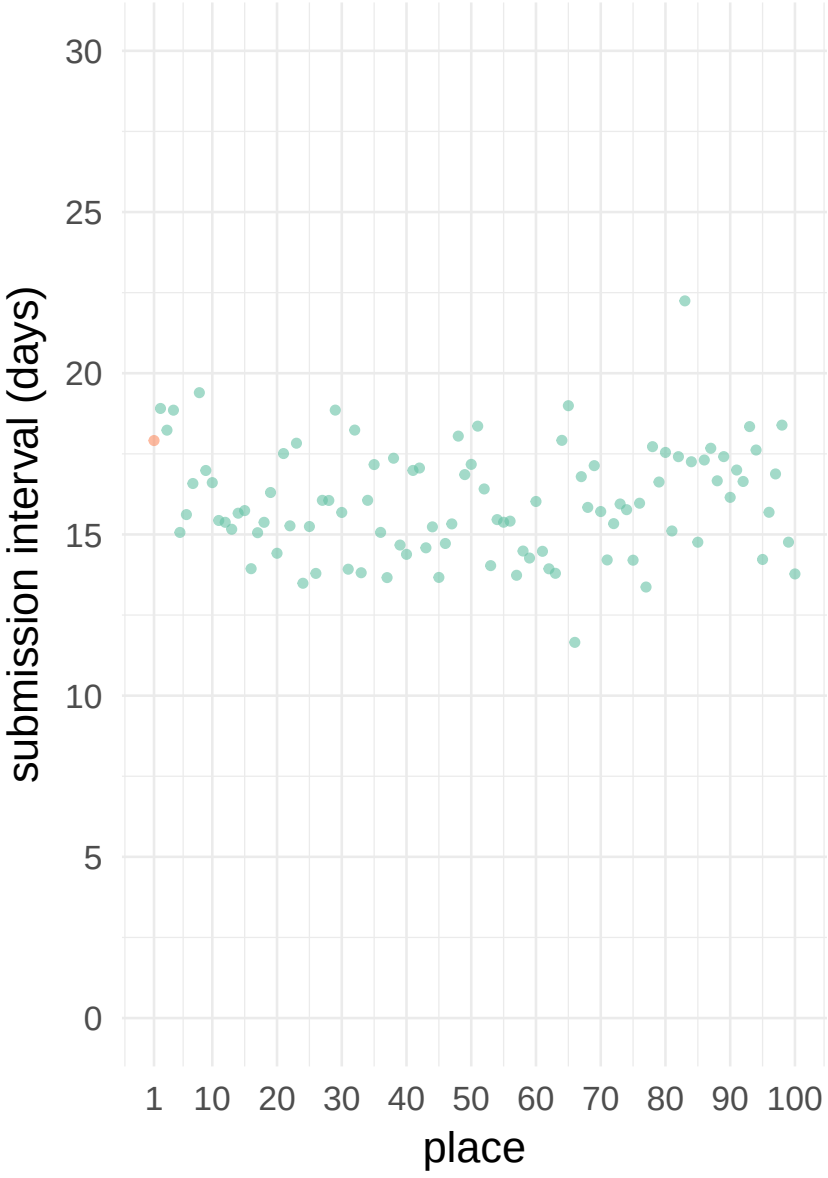


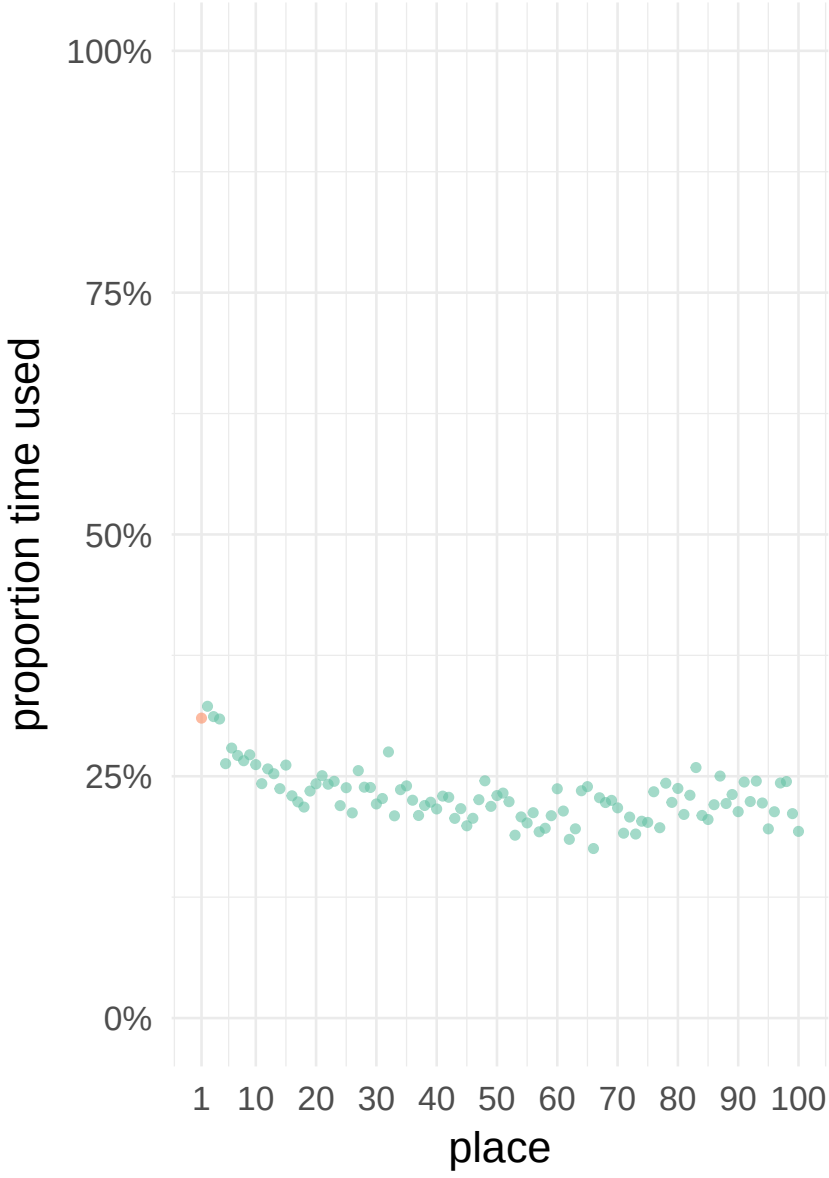




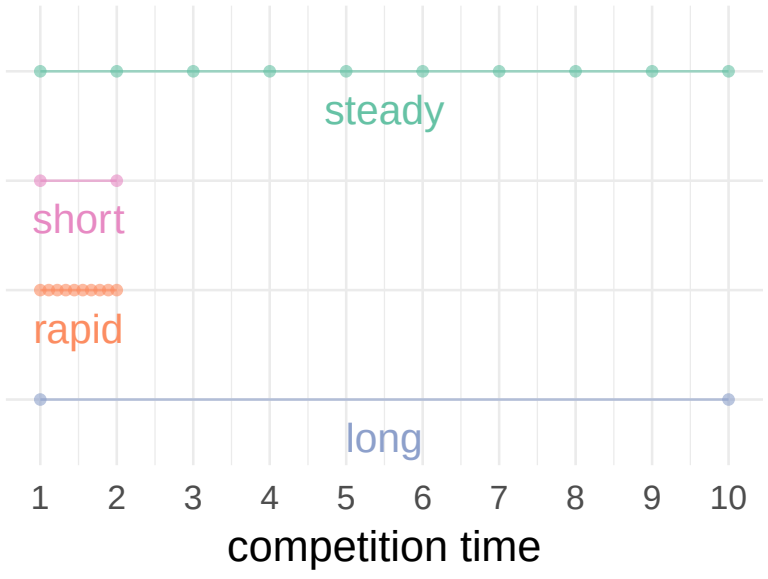


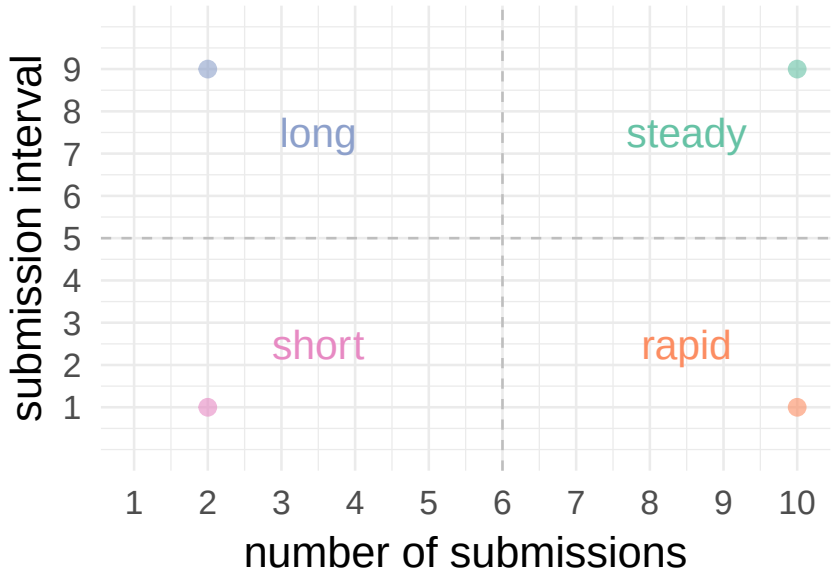
Submission interval

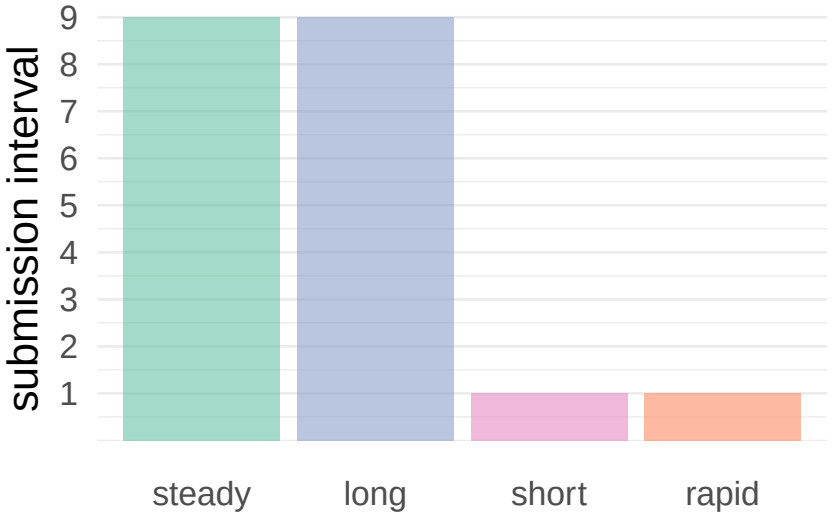


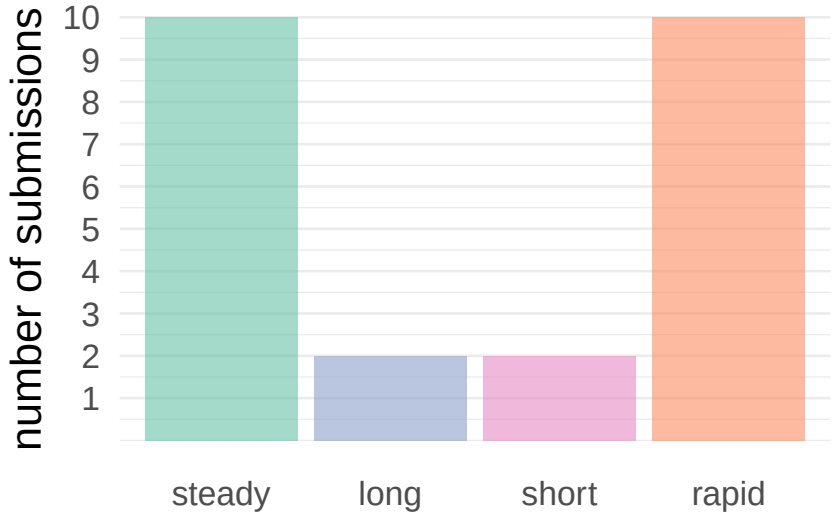


Types of strategies

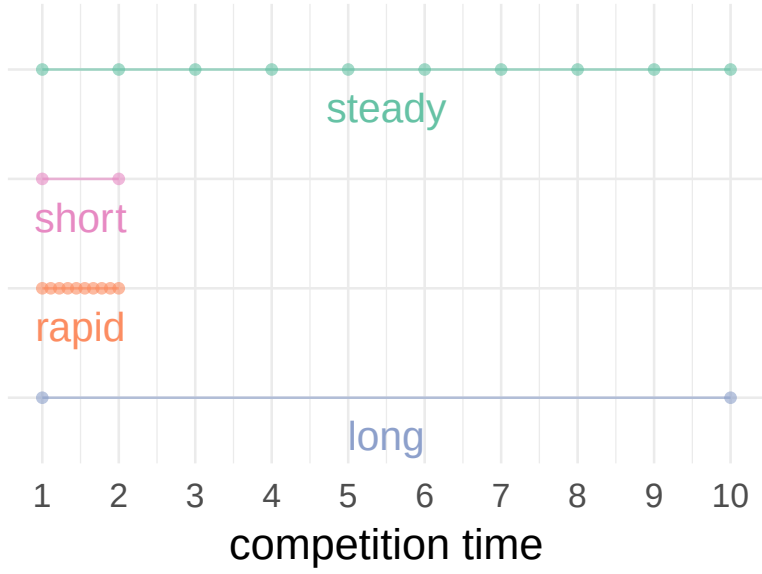


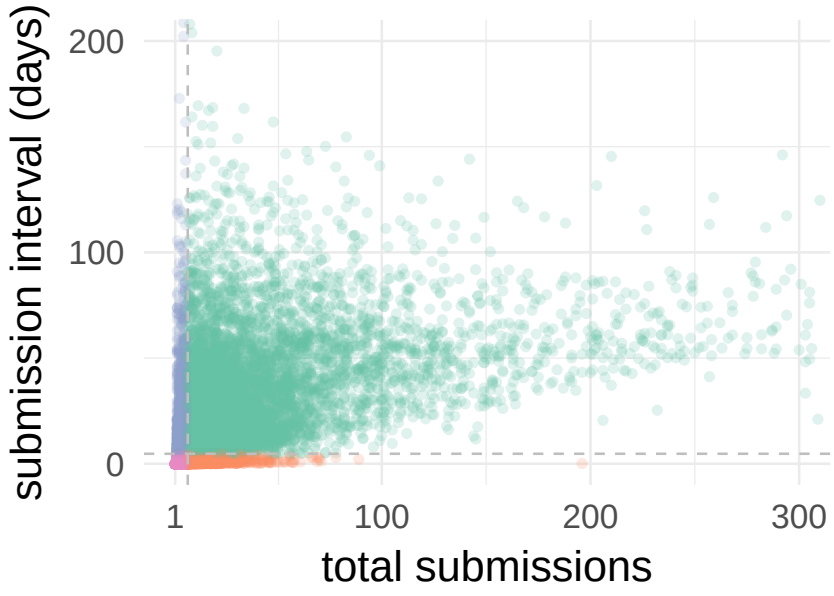


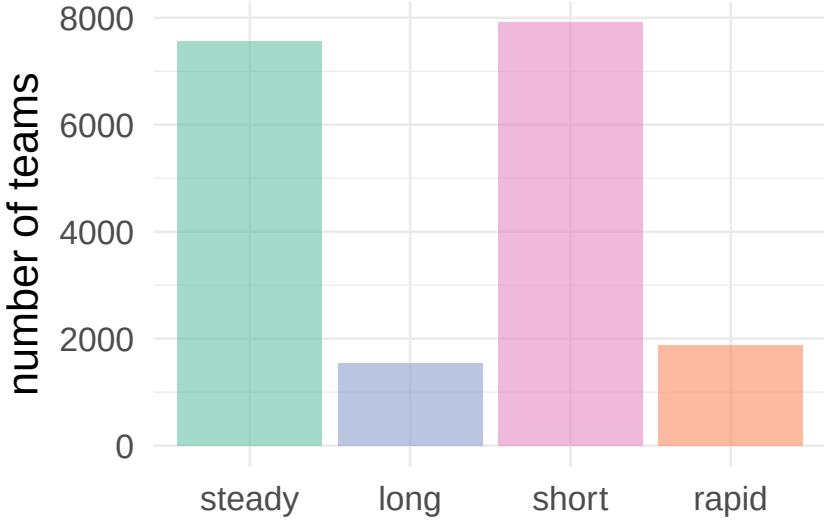


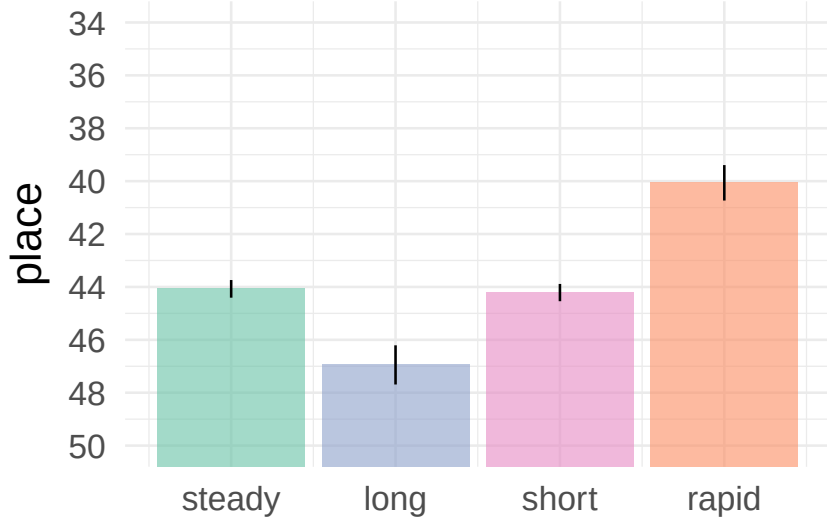


Iteration as a strategy









Is Wikipedia getting better?

My favorite way of checking this is to "click random article" on 10 articles, and go back and look at them a year ago, 5 years ago, 10 years ago. Every time I have tried, it's unambiguous: Wikipedia is getting better by this test. – Jimbo Wales

Wikipedia by the numbers

- 5 million articles in English.
- 5th most popular website in the world.
- 6-10 edits per second.
- 700 new articles per day.

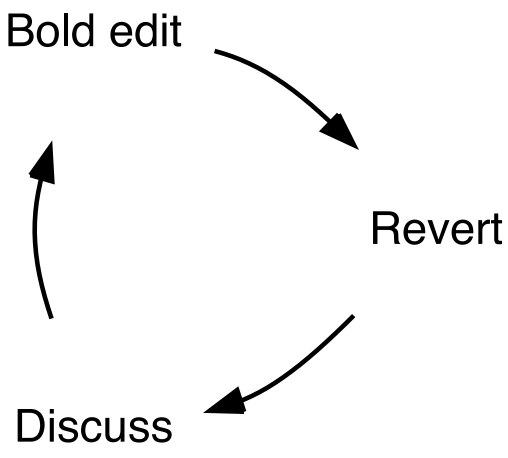
Wikipedians

- 140,000 active users ($<$ 30 days).
- Vandalism detection is highly automated.
- New editors do not like getting reverted.
- All editors are protective of their own edits.
- Chance of being reverted doesn't change.

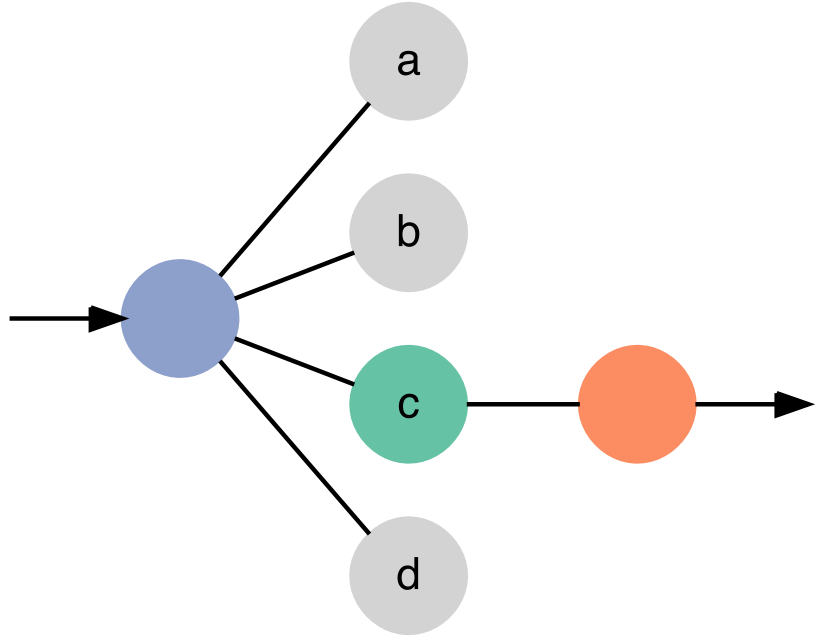
(For more, start with Aaron Halfaker).

Wikipedia article editing as an evolutionary strategy

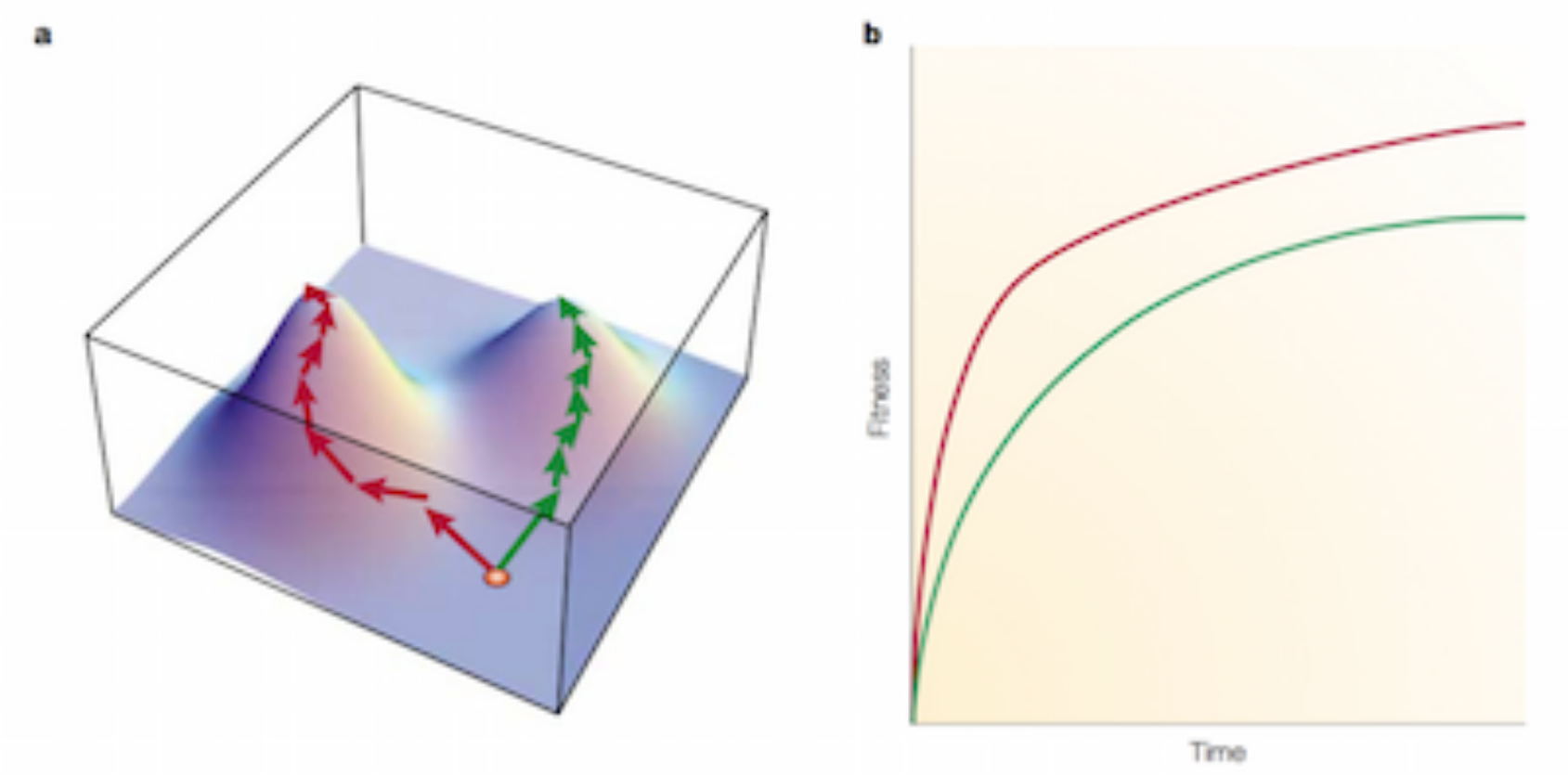
The BRD cycle editing strategy



Results of BRD cycle editing



Experimental evolution (Elena & Lenski, 2003)

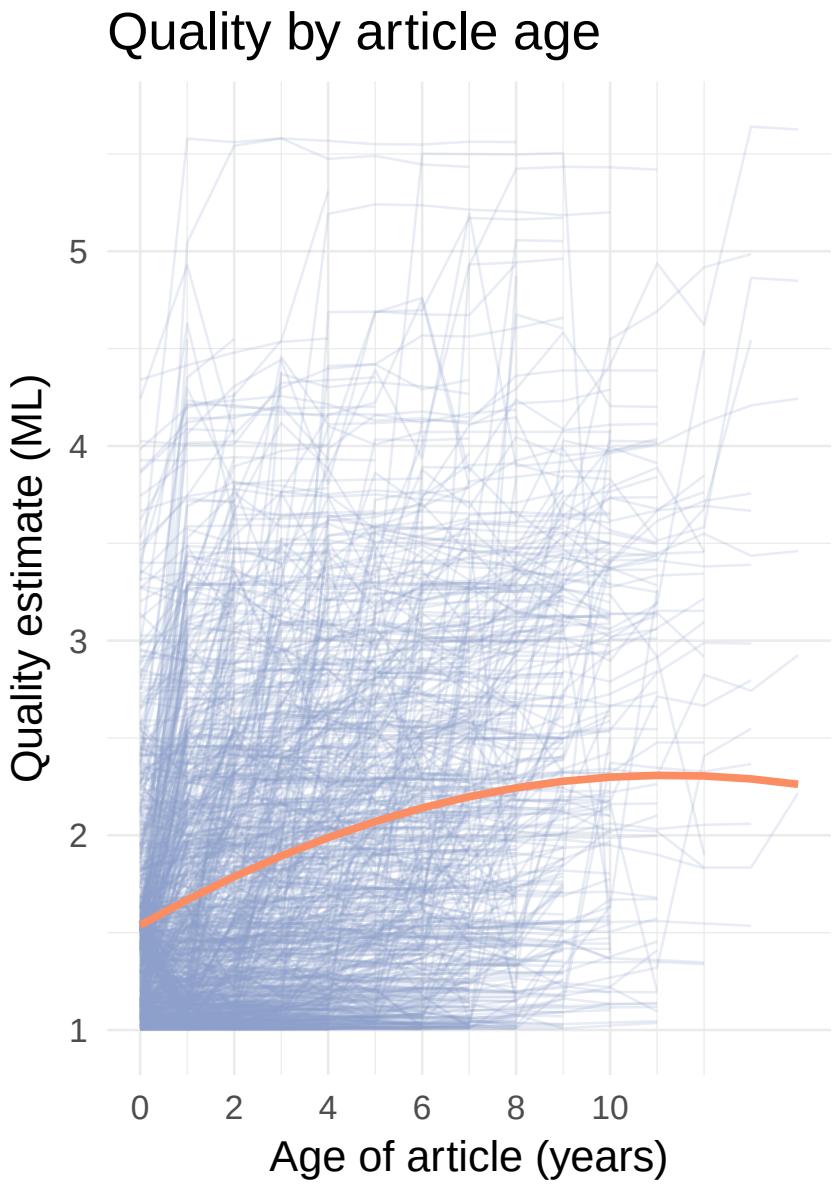


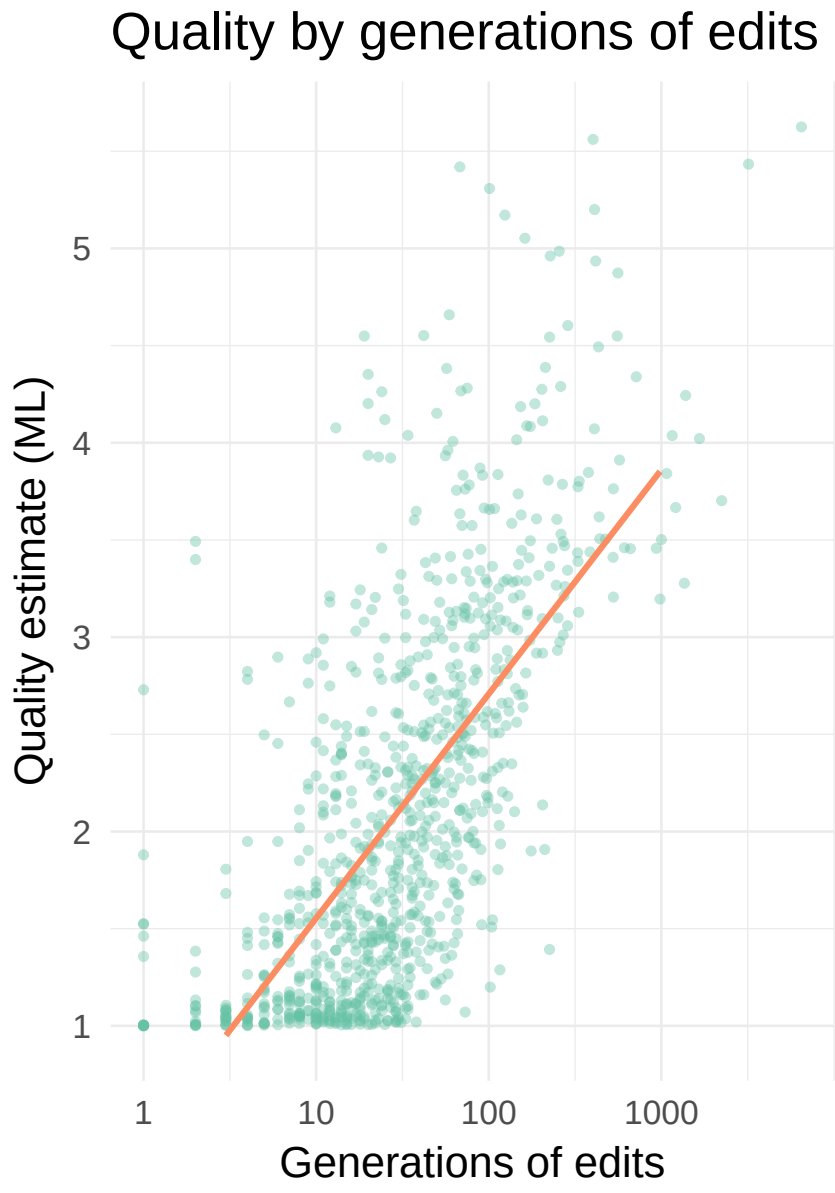
Predicting article quality (Warncke-Wang et al., 2015)

All rated articles by quality and importance						
Quality	Importance					
	Top	High	Mid	Low	???	Total
FA	1,109	1,709	1,617	966	179	**5,580**
FL	141	540	634	578	113	**2,006**
A	194	391	546	327	71	**1,529**
GA	1,881	4,328	8,491	8,698	1,608	**25,006**
B	11,376	21,728	32,767	25,146	13,305	**104,322**
C	9,210	26,613	58,912	76,564	39,438	**210,737**
Start	16,396	70,926	284,777	691,016	270,191	**1,333,306**
Stub	4,213	29,490	211,014	1,684,123	827,029	**2,755,869**
List	2,774	10,192	29,954	80,224	58,075	**181,219**
Assessed	47,294	165,917	628,712	2,567,642	1,210,009	**4,619,574**
Unassessed	115	344	1,614	16,840	487,999	**506,912**
Total	**47,409**	**166,261**	**630,326**	**2,584,482**	**1,698,008**	**5,126,486**

Authority/Reputation, **Completeness**, Complexity, **Informativeness**, Consistency, Currency, Volatility, Diversity, **NumHeadings**, **ArticleLength**, **NumReferences**, NumWikilinks, HasInfobox, . . .

Monotonic increases in Wikipedia article quality





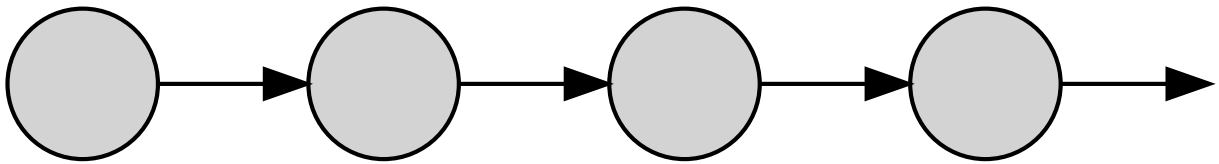
Future directions

- Edit quality models (big data!).
- Separate purifying from positive selection.
- Expand to open source software projects.

Discussion

What do you think of the Wikipedia/evolution comparison?

Summary

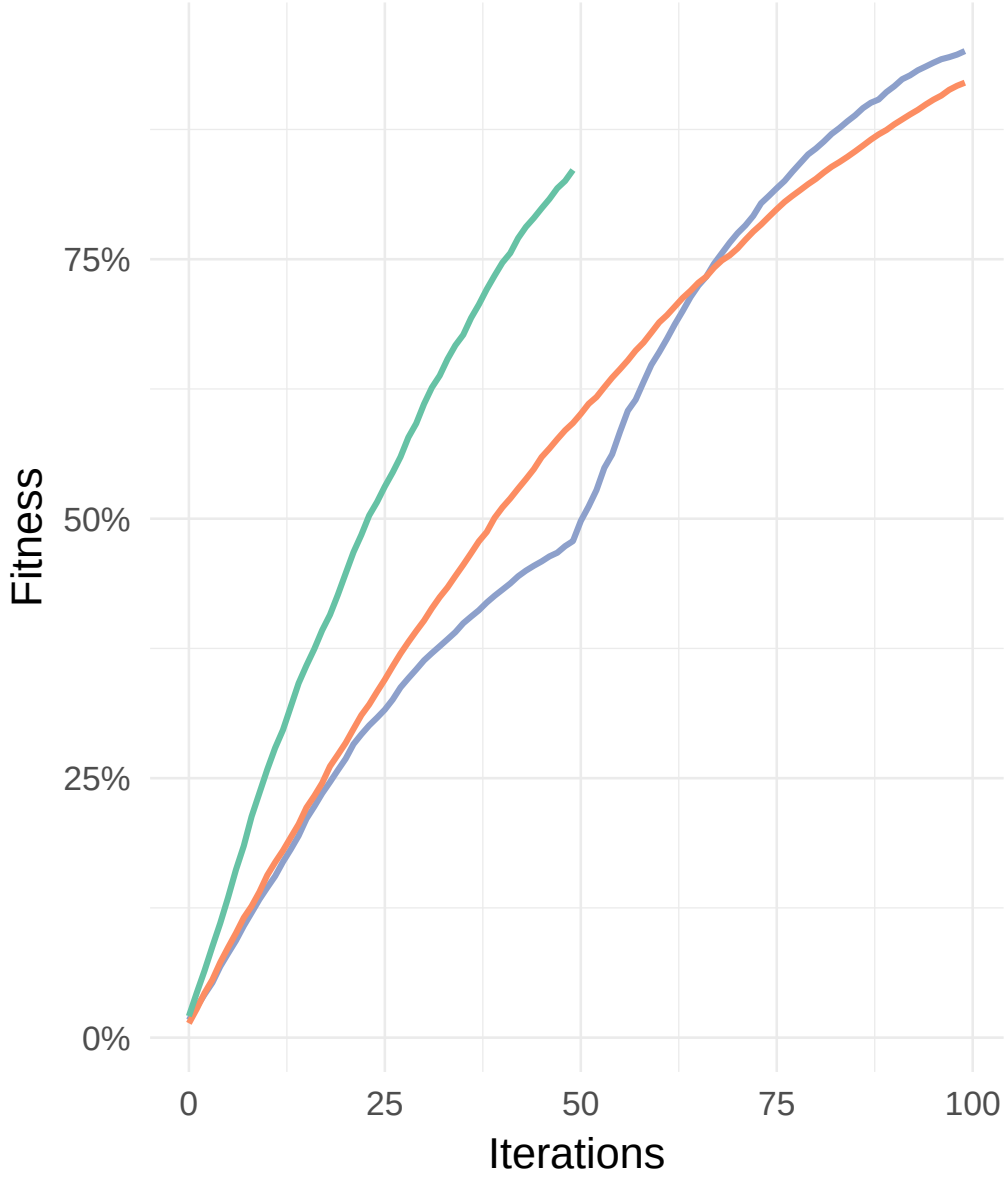


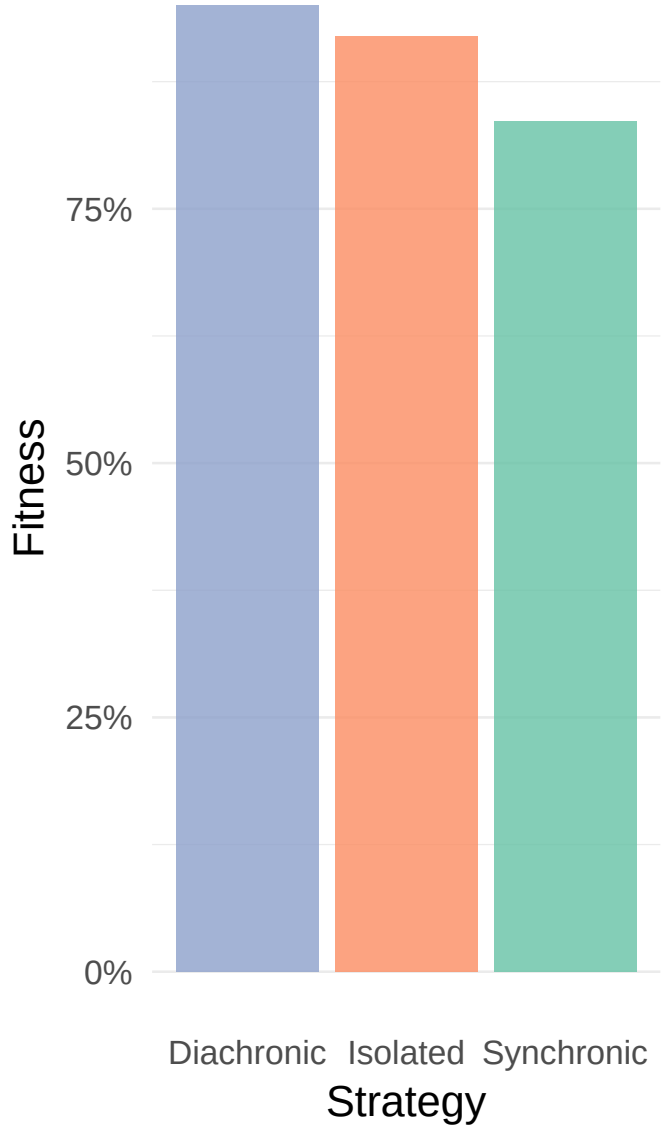
- Word origins: Imitations transition to words through repetition
- Problem solving: Diachronic inheritance, Iteration as a strategy
- Wikipedia article editing as an evolutionary strategy

Using evolutionary principles to make sense of
word origins,
problem solving, and
the growth of Wikipedia articles.

Pierce Edmiston
pedmiston@wisc.edu
github.com/pedmiston/leaning-on-darwin

Proof of principle





Problem solving as hill climbing