# Teaching an old dog new tricks? Learning rates, aging, and language change

Ellis Cain

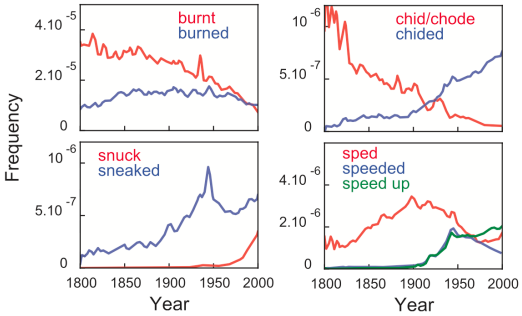
May 2, 2023

Section 1

# Background literature

# Collective patterns of language usage

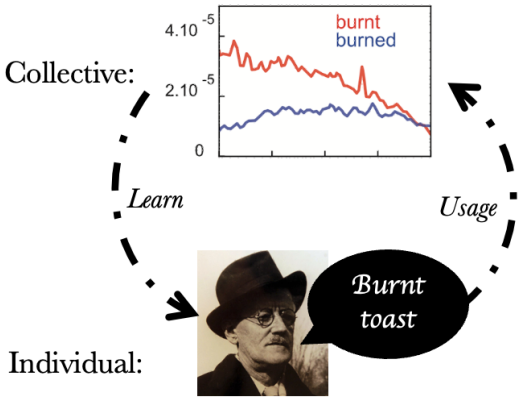
- N-gram corpus analysis of Google Books corpus[1]
- Diachronic changes in language usage and meaning[2]




[1] (Michel et al., 2011)
[2] (Bybee, 2015; Bynon et al., 1977)

# Collective patterns of language

*Emergence of collective-level trends from individual usage*



# Mechanisms of language acquisition

- Statistical learning[3]
- Propose but verify (hypothesis testing)[4]
- Structural inference[5]

[3] (Smith & Yu, 2008; Yu & Smith, 2007)
[4] (Trueswell et al., 2013)
[5] (Kim et al., 2019)

# Aging and learning

- Initially rely on associative / bottom-up learning
- Later shift to inferential / top-down learning

Section 2

# Formal model

# Overview

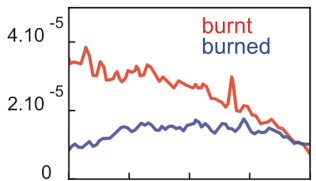
- Language change as interaction between individual and collective level dynamics
- Aim to explore how individual learning rates, aging, and group membership impact overall population-level patterns of language change

[6](Troutman et al., 2008)

# Overview

- Language change as interaction between individual and collective level dynamics
- Aim to explore how individual learning rates, aging, and group membership impact overall population-level patterns of language change

- Model of the usage and spread of a grammatical variant throughout a population[6]
  - Past tense ending can be "-t" or "-ed", such as in "burnt" or "burned"




[6](Troutman et al., 2008)

# Model assumptions[7]

1. Language learning is based on imitating others
   - E.g., individuals may learn quickly early on, but slow down as they age

# Model assumptions[7]

1. Language learning is based on imitating others
   - E.g., individuals may learn quickly early on, but slow down as they age
2. There are variations in preference between individuals
   - E.g., some individuals learn more quickly than others

# Model assumptions[7]

1. Language learning is based on imitating others
   - E.g., individuals may learn quickly early on, but slow down as they age
2. There are variations in preference between individuals
   - E.g., some individuals learn more quickly than others
3. Language can be influenced by external factors
   - E.g., more willing to learn from in-group members

[7] (Beeksma et al., 2017)

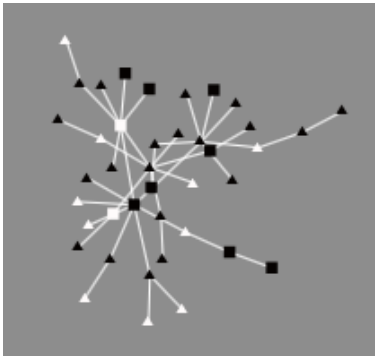
# Model assumptions[7]

1. Language learning is based on imitating others
   - E.g., individuals may learn quickly early on, but slow down as they age
2. There are variations in preference between individuals
   - E.g., some individuals learn more quickly than others
3. Language can be influenced by external factors
   - E.g., more willing to learn from in-group members
4. Language change has multiple stable equilibriums

[7] (Beeksma et al., 2017)

# Initialization: network

- Generates preferential attachment network
- Distributes grammar according to specified percentage of grammar 1
  - Two grammar variants, 0 or 1 (*burnt or burned*)



# Initialization: nodes

Represent language users

- State: node's current grammar preference, initialized as 0 or 1
- Age
  - Probablistic or deterministic
- Cohort: "Age group", *either 1 or 2*
- Gamma: learning rate of a given node
  - Probablistic, deterministic, or based on age

# Initialization: cohorts

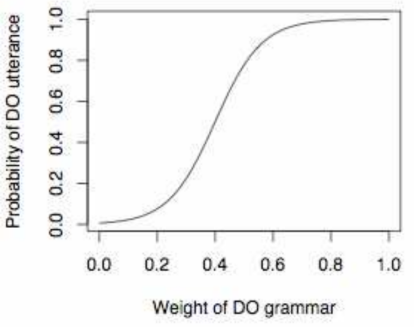
- Number of cohorts (max 2) based on specified percentage
- Cohort ages can be specified
- Option for cohort-based grammar, such that the cohorts start with different percentages of grammar 1
- Willingness to listen to out-group members

# Dynamics

- Communication
  - Speaking (asynchronous)
  - Neighboring agents listen
- Aging

# Dynamics: speaking

- Nodes will generate an 'utterance', which is either 0 or 1 (*burnt or burned*)
- Nodes 'prefer' a discrete grammar
- Logistic curve is used when nodes produce an utterance



# Dynamics: listening

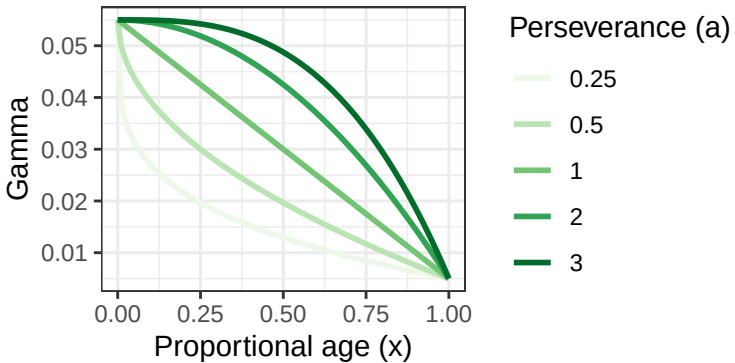
- Neighboring nodes will pick a grammar that will be used to interpret heard utterance
- If it matches the heard utterance
  - Update listener's state towards the heard state
  - Otherwise, it will update listener's state away from the heard state
  - Learning rate: *Gamma parameter modifies the step size*
- *Chance to ignore out-group*

# Dynamics: Aging

- Nodes age with each tick
- Gamma changes with age: either constant or decreasing with age

# Dynamics: Aging

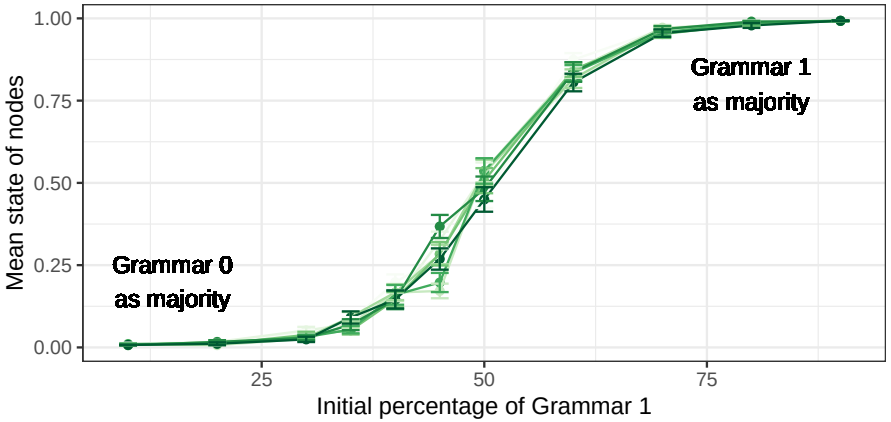
- Nodes age with each tick
- Gamma changes with age: either constant or decreasing with age
- Perseverance: how slowly gamma decays
  - Basic power law: $y = -0.05(x^a) + 0.005$



Section 3

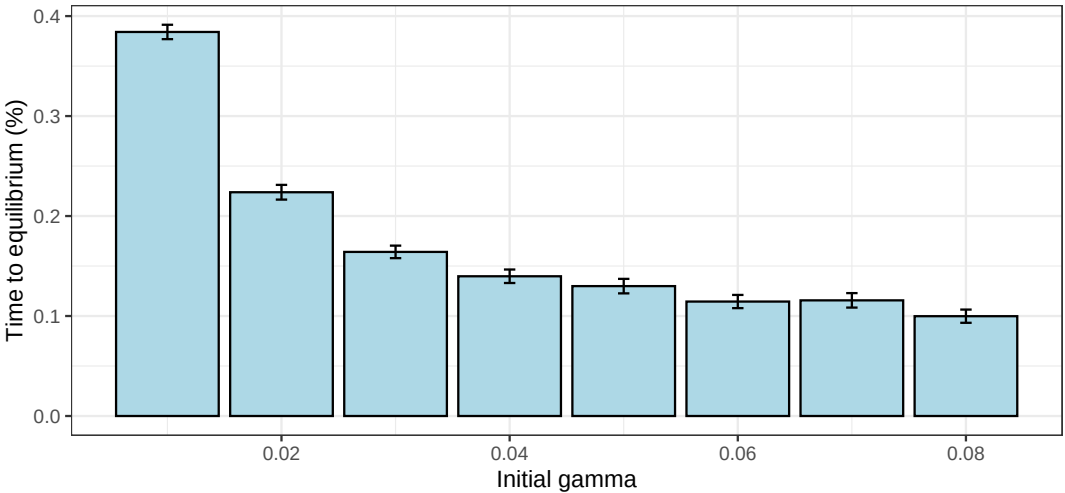
# Results

# Impact of learning rate




50 runs per configuration, 1000 steps total per run

# Impact of learning rate

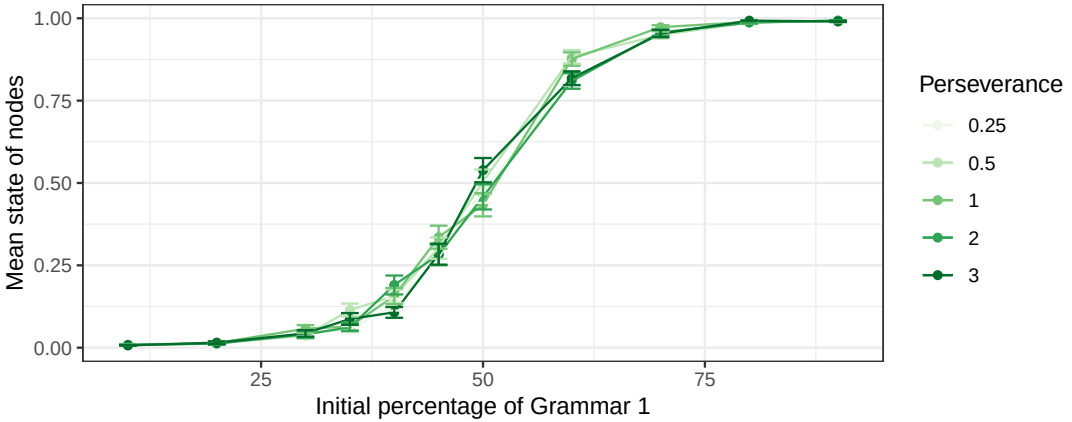



50 runs per configuration, 1000 steps total per run

# Checkpoint

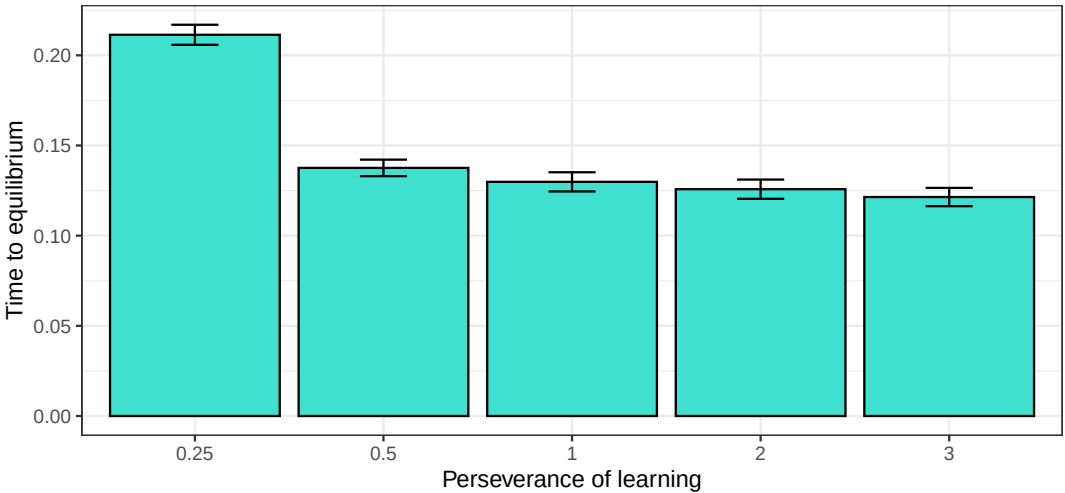
- Increased gamma decrease time to equilibrium, but not (systematically) affect the final outcome

# Variation of speed of decrease (perseverance)




50 runs per configuration, 1000 steps total per run

# Variation of speed of decrease (perseverance)

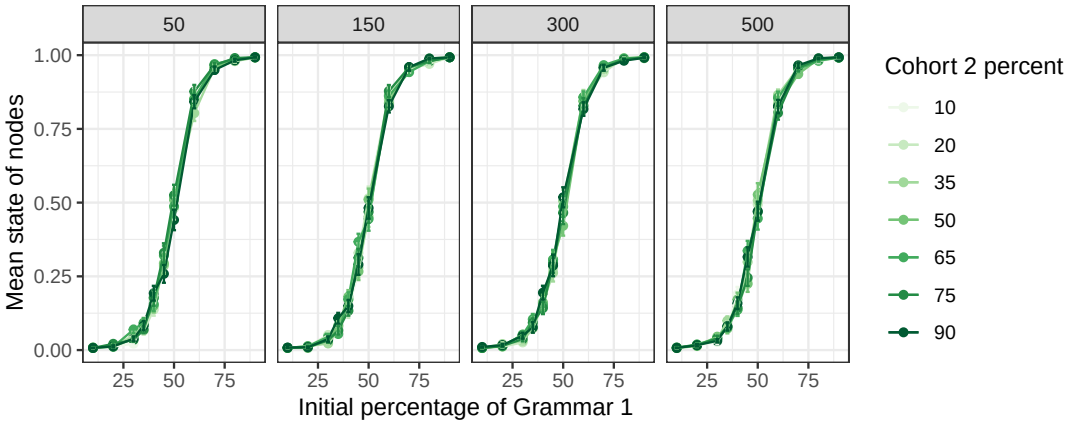



50 runs per configuration, 1000 steps total per run

# Checkpoint

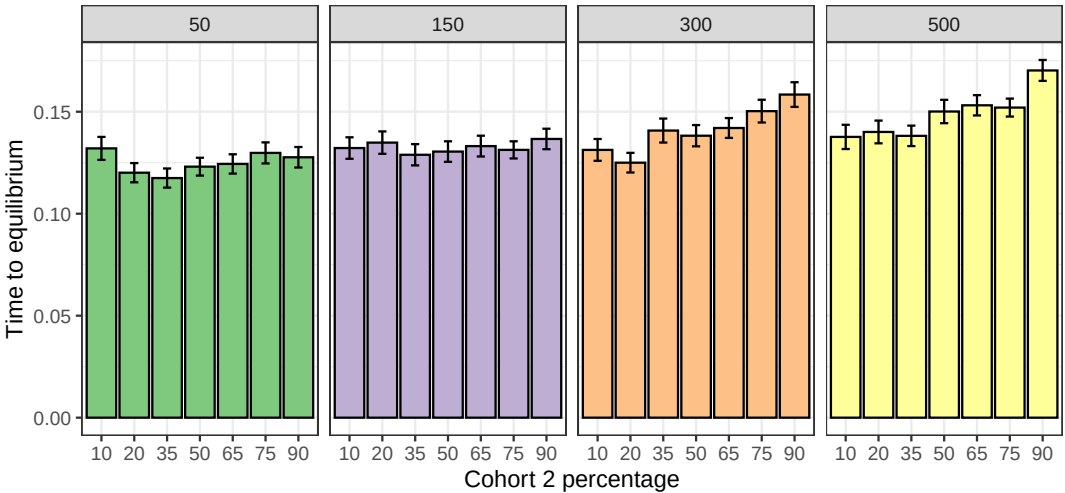
- Gamma impacts time to equilibrium (TTE)
- No difference in TTE when perseverance is $> 0.5$

# Two age cohorts



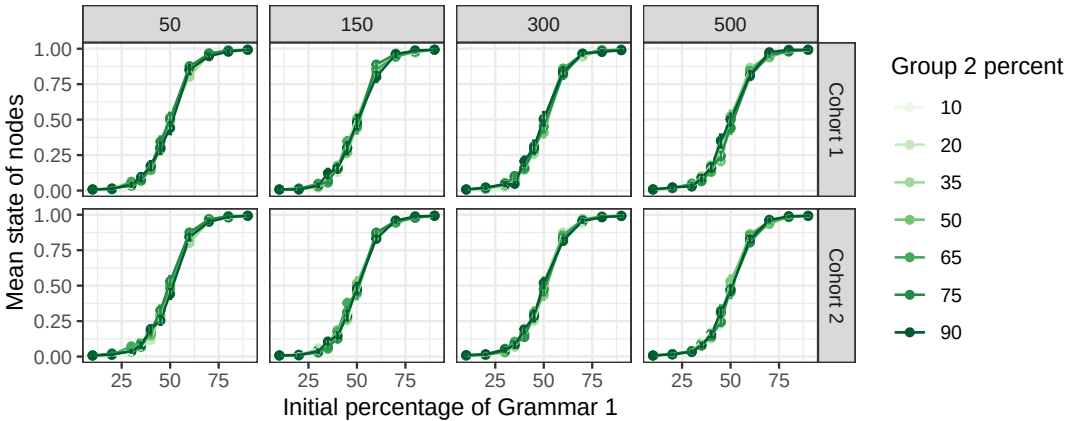

Faceted by Cohort 2 age; 50 runs per configuration, 1000 steps total per run

# Two age cohorts



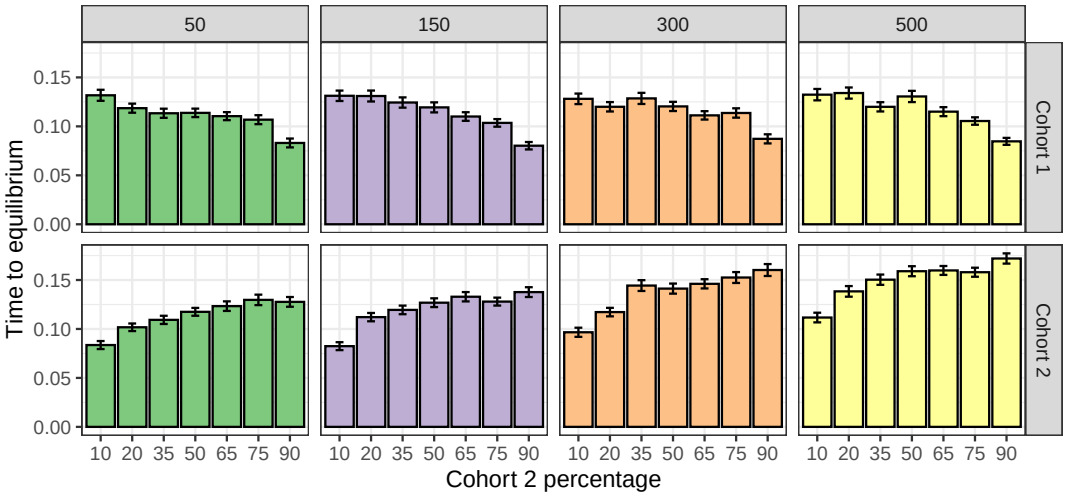

Faceted by cohort 2 age; 50 runs per configuration, 1000 steps total per run

# Two age cohorts: group equilibria




50 runs per configuration, 1000 steps total per run

# Two age cohorts: group equilibria

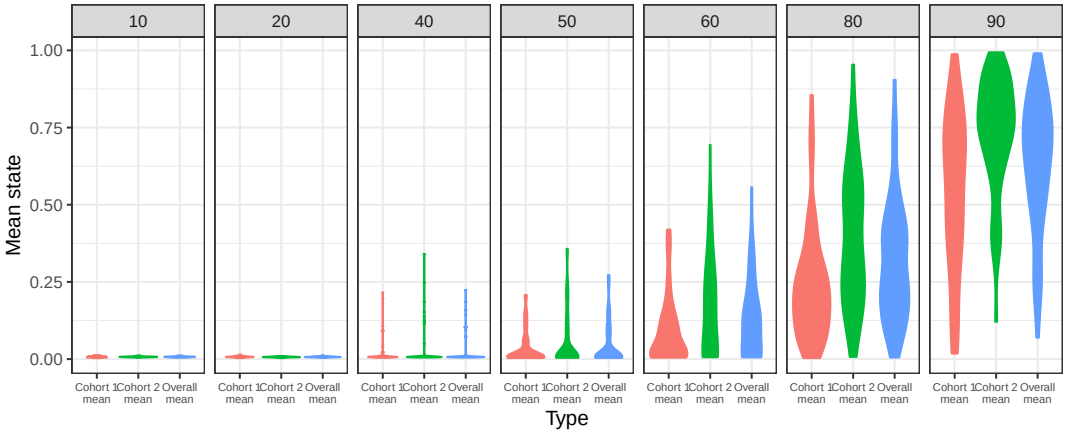



Faceted by cohort 2 age; 50 runs per configuration, 1000 steps total per run

# Checkpoint

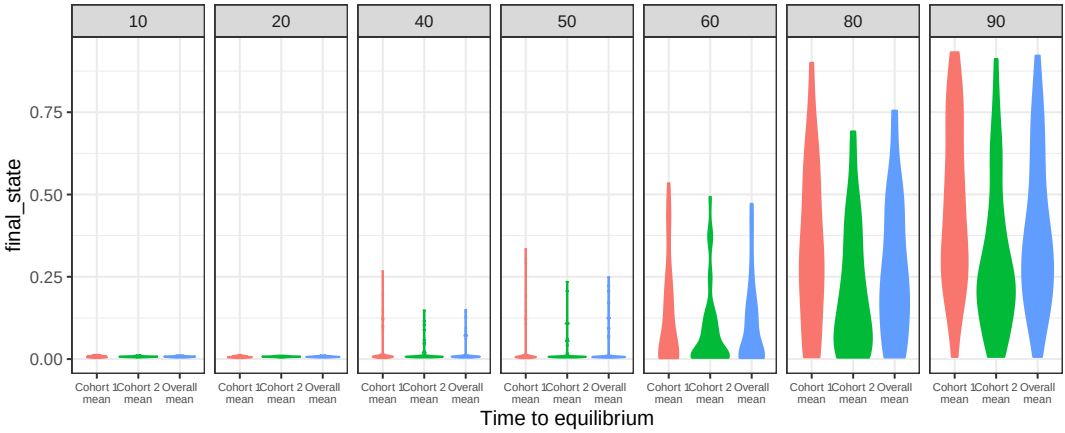
- Gamma impacts time to equilibrium (TTE)
- No difference in TTE when perseverance is $> 0.5$
- Older cohort slows down TTE
- As the age gap increases, the impact on TTE increases as well

# Cohort-based grammar: Only Cohort 2 has grammar 1




Gamma decreases at constant rate; Cohort 1 does not have grammar 1; 50% Cohort 2; 50 runs per configuration, 1000 steps total per run

# Cohort-based grammar: Only Cohort 1 has grammar 1

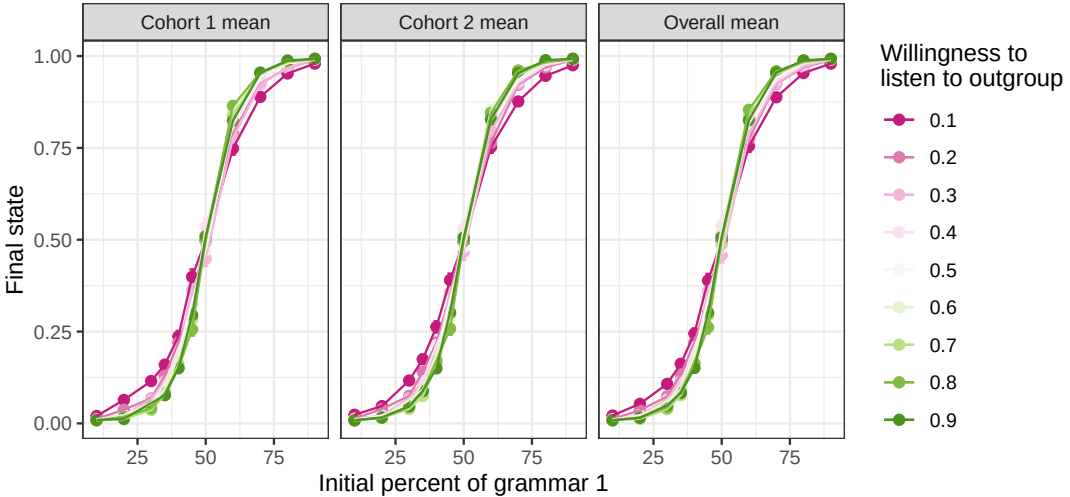



Gamma decreases at constant rate; Cohort 2 does not have grammar 1; 50% Cohort 2; 50 runs per configuration, 1000 steps total per run

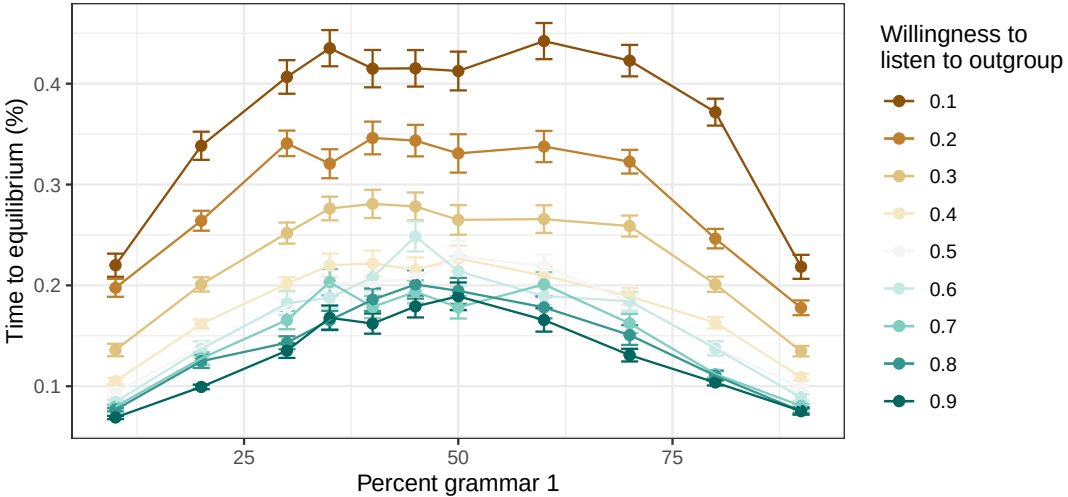
# Checkpoint

- Gamma impacts time to equilibrium (TTE)
- No difference in TTE when perseverance is $> 0.5$
- Older cohort slows down TTE, exacerbated by increased age gap
- Reservoir? When grammar varies only in the older cohort, their initial starting percentage will draw the younger cohort
- Innovator: when grammar varies only in the younger cohort, However, when cohort 2 does not use grammar 1, cohort 1 starts with increasingly higher percentages of using grammar 1, they themselves may end up with using grammar 1 more, but it does not drive the overall group usage.

# Cohort preference



# Cohort preference



# Checkpoint

- Increased gamma speeds up when equilibrium is reached, but not the final outcome; modulated by perseverance
- Older cohort delays equilibrium
- “Innovator” and “Reservoir” groups
- Group preference will delay the equilibrium, and slightly impact equilibrium value

# References

Beeksma, M., Vos, H. de, Claassen, T., Dijkstra, A., & Kemenade, A. van. (2017). *A probabilistic agent-based simulation for community level language change in different scenarios*.

Bybee, J. (2015). *Language change*. Cambridge University Press.

Bynon, T. et al. (1977). *Historical linguistics*. Cambridge University Press.

Kim, J. S., Elli, G. V., & Bedny, M. (2019). Knowledge of animal appearance among sighted and blind adults. *Proceedings of the National Academy of Sciences*, *116*(23), 11213–11222.

Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., The Google Books Team, Pickett, J. P., Hoiberg, D., Clancy, D., Norvig, P., Orwant, J., Pinker, S., Nowak, M. A., & Aiden, E. L. (2011). Quantitative Analysis of Culture Using Millions of Digitized Books. *Science*, *331*(6014), 176–182. https://doi.org/10.1126/science.1199644

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558–1568.

Troutman, C., Clark, B., & Goldrick, M. (2008). Social networks and intraspeaker