# LI 308: Computational Models of Sound Change

Day 4: Phonetic change in a population setting

James Kirby & Morgan Sonderegger
30 July, 2015

# Introduction

- Change at the population level is often claimed to be based in phonetic variation at the individual level (e.g. Ohala, 1993)
- One source[1] of variation: production bias (e.g., coarticulation)

| *WGmc* | *Pre-OHG* | *OHG (NHD)* |
|---|---|---|
| *gasti | gesti | gest (*Gäste)* |
| *lambir | lembir | lemb (*Lämme)* |
| *fasti | festi | fest (*fest)* |

Primary umlaut in West Germanic (after Iverson and Salmons, 2006).

[1] (But certainly not the only one! also group membership, cognitive endowment...)

# From individuals to populations

- '...the initiation of such sound changes is accomplished by the phonetic mechanism just described; their spread, however, is done by social means, e.g., borrowing, imitation, etc.' (Ohala, 1981:184)

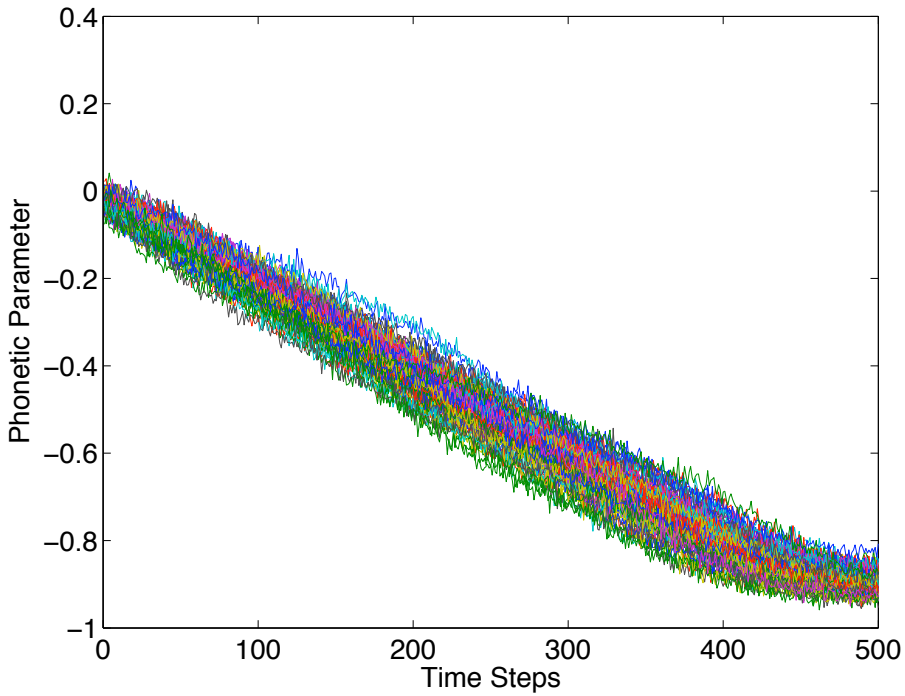
# From individuals to populations

- '...the initiation of such sound changes is accomplished by the phonetic mechanism just described; their spread, however, is done by social means, e.g., borrowing, imitation, etc.' (Ohala, 1981:184)

- 'I will try to show how a change in the pronunciation norm of a given word occurs in at least one speaker; what happens to this changed norm after that will involve different mechanisms and ones not properly part of sound change.' (Ohala, 1989:175)

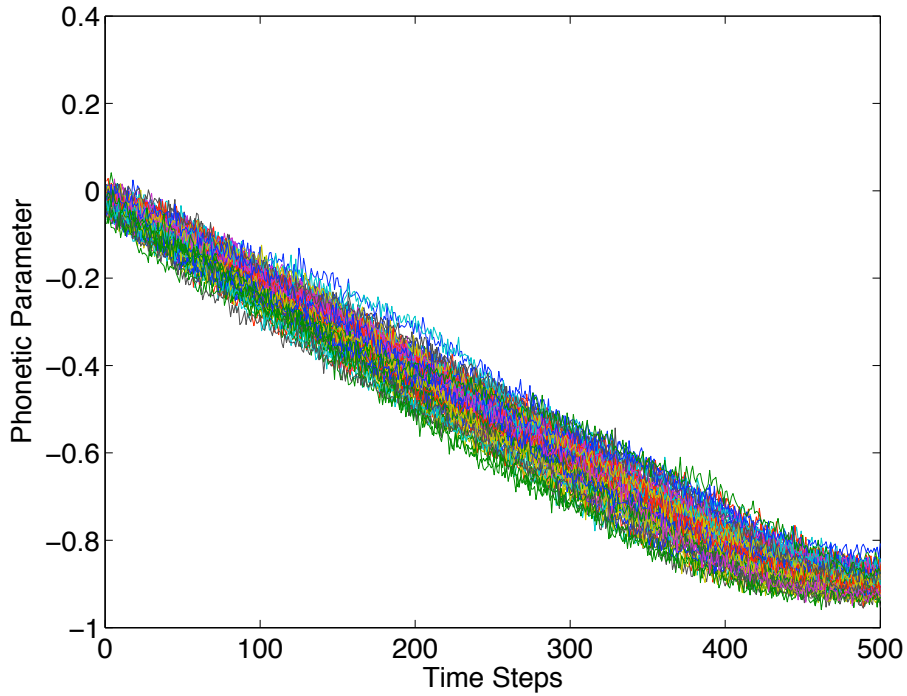
# Stability and change

- Change in individuals is a *necessary* but not *sufficient* condition for change at the population level



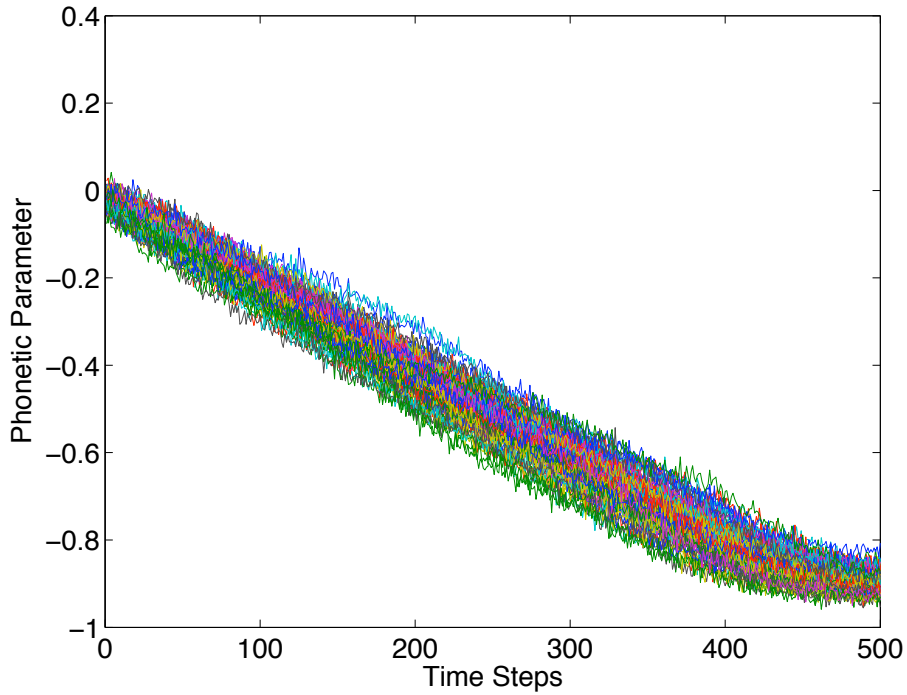
# Stability and change

- "Accumulation-of-error" approaches often criticized for this very reason (e.g. Baker, 2008)



# Stability and change

- For one thing, *existence* of a bias does not mean change is inevitable: default is stability! (Weinrich et al., 1968; cf. Kiparsky's "non-phonologization problem")

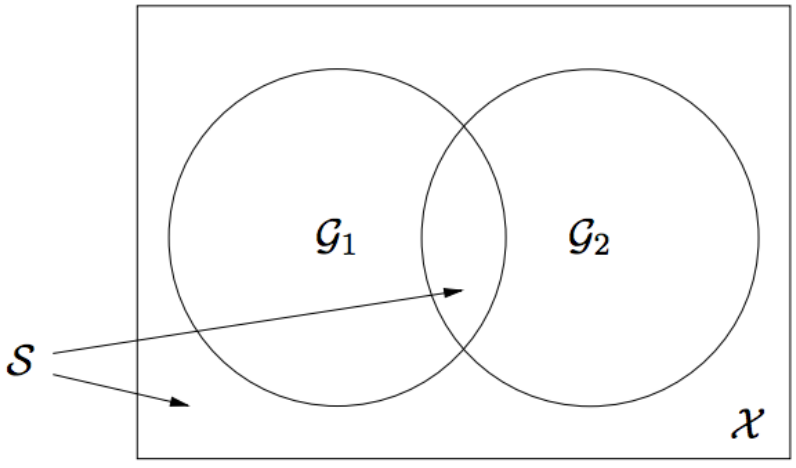


# Stability and change

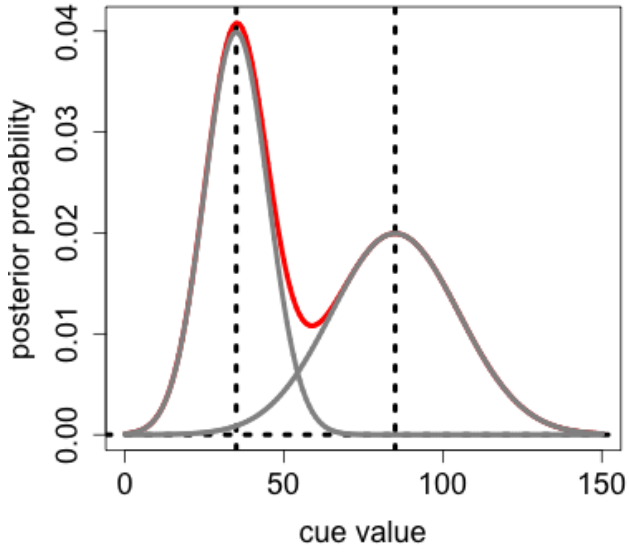
- Thus, an adequate account of actuation must explain:
  1. Stability of limited coarticulation in the population;
  2. Stability of full coarticulation in the population;
  3. Change from stable limited to full coarticulation.

# Continuous vs. discrete

- Most computational work on language evolution has focused on learning discrete parameters (Niyogi & Berwick, 1995; Griffiths & Kalish, 2007; S. Kirby et al., 2007...)




(Griffiths & Kalish, 2007)




- However, phonetic learning also involves continuous parameters (Maye et al., 2002; Clayards et al., 2008; Feldman et al., 2013...)

# Bias and population structure

- Change in individuals is a necessary but not sufficient condition for change at the population level
- Both learning and population structure are important
  - Different assumptions about learning $\rightarrow$ different outcomes (Kirby, 2002; Zuidema, 2003; Brighton et al., 2005; Kirby et al., 2007)
  - Most modelers assume fixed population structure (e.g. diffusion chains) but dynamics for arbitrary populations are in general nonlinear (Niyogi & Berwick 1995; Dediu, 2009; Smith, 2009)
- What role do these play in the continuous learning case?

# Roadmap

- Today: framework in which stability and change at the population level is possible be assuming both
  1. a force promoting contrast maintenance, to keep separate phonetic categories stable; and,
  2. an external force, such as a production bias, which induces change (cf. Pierrehumbert, 2001; Wedel, 2006).

Kirby & Sonderegger (2015)

# Roadmap

- Then: some new questions
  1. Does using production bias as the external force have a unique dynamics?
  2. If not, will *any* kind of external force produce the same behaviour at the population level?

# Roadmap

- Our example scenario: phonologization of coarticulation

| *WGmc* | *Pre-OHG* | *OHG (NHD)* |
|---|---|---|
| *gasti | gesti | gest (*Gäste)* |
| *lambir | lembir | lemb (*Lämme)* |
| *fasti | festi | fest (*fest)* |

# Roadmap

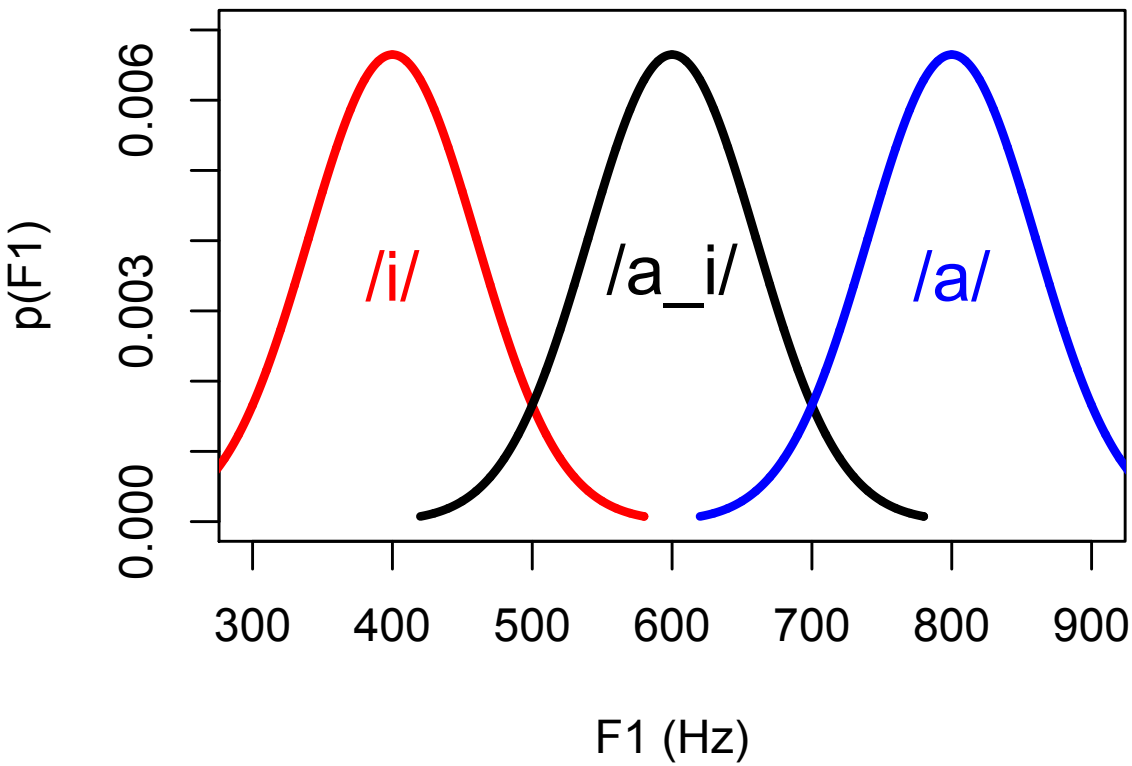
- Our example scenario: phonologization of coarticulation

| *WGmc* | *Pre-OHG* | *OHG (NHD)* |
|---|---|---|
| *gasti | gesti | gest (*Gäste)* |
| *lambir | lembir | lemb (*Lämme)* |
| *fasti | festi | fest (*fest)* |

- Simple models, iterated over generations $\Rightarrow$ potentially unintuitive outcomes

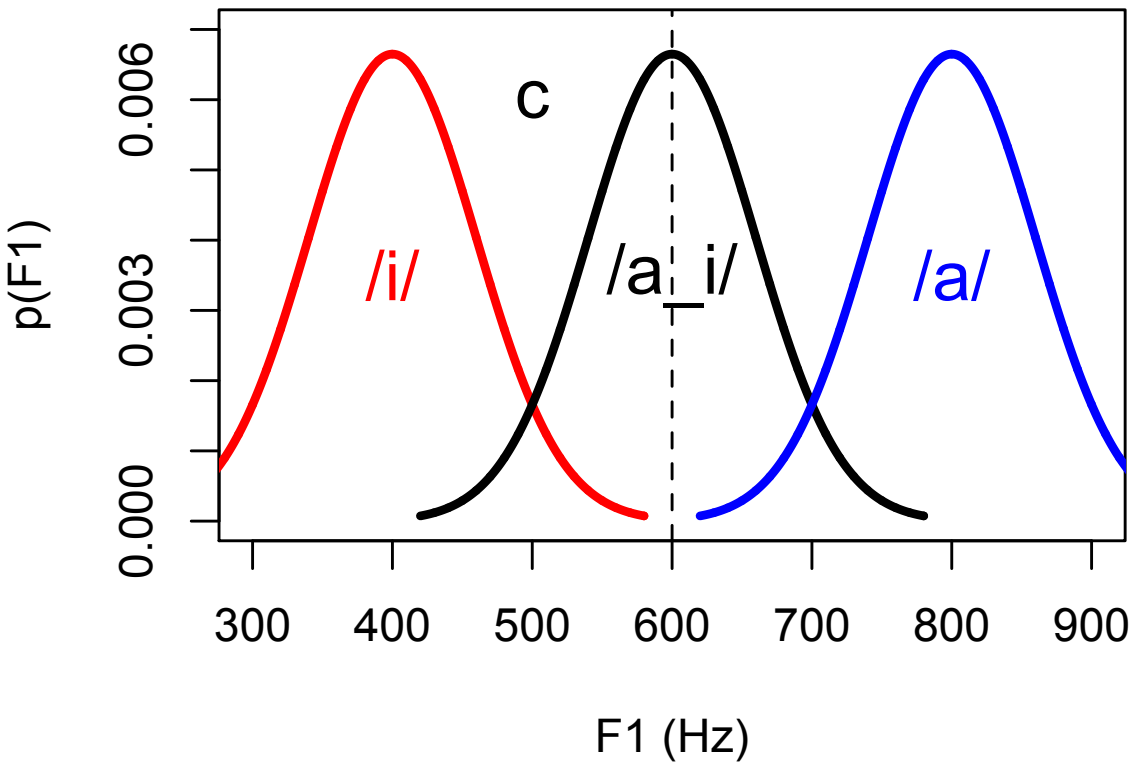
# Framework

- *Lexicon*: $\{V_1, V_2, V_{12}\}$, where $V_{12}$ represents $V_1$ in the coarticulation-inducing context of $V_2$



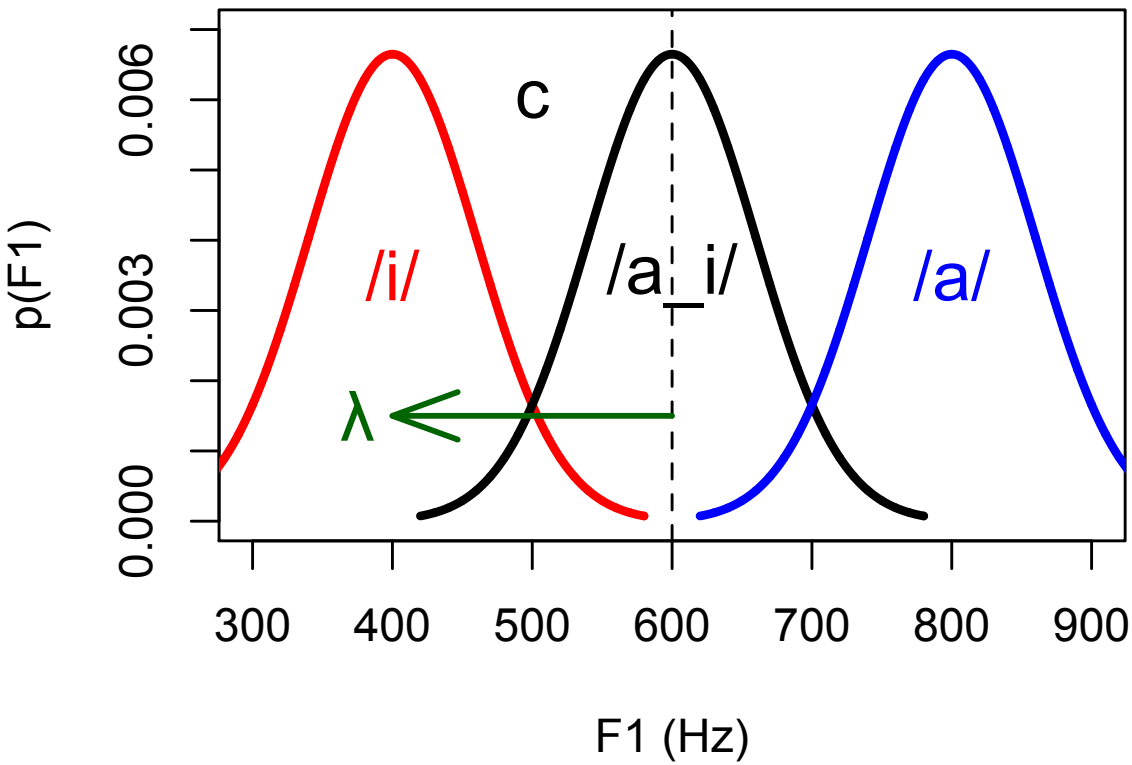
# Framework

- *Task:* learn a **contextual variant** *c*: how much /a/ is produced like /i/ in the context of /i/ (/a_i/)



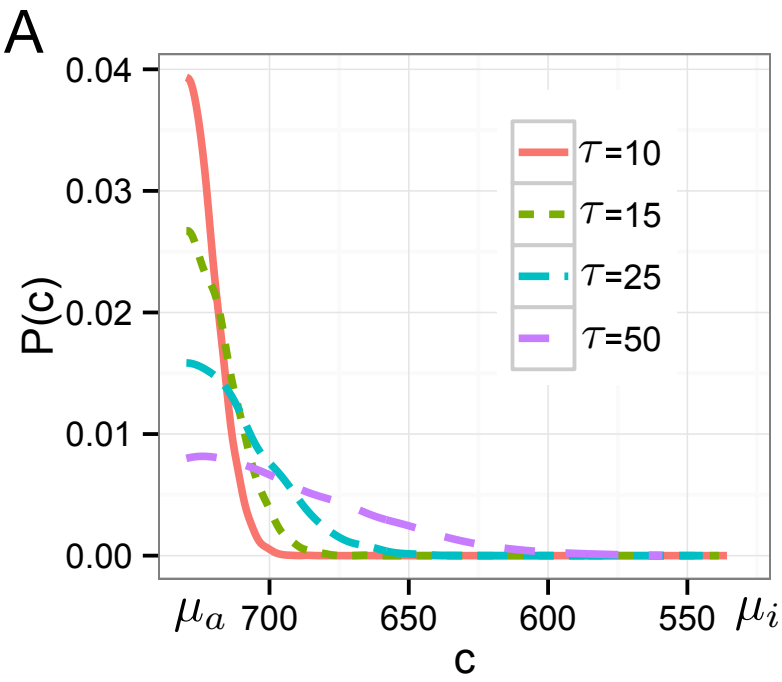
# Framework

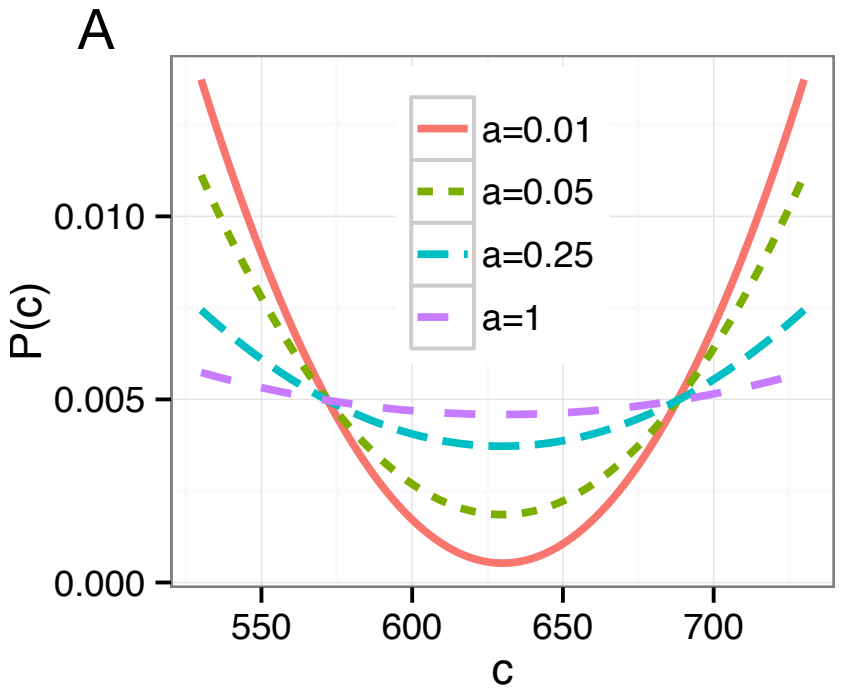
- *Data:* F1 values for /a_i/ tokens, potentially subject to **production bias** $\lambda$ (assuming fixed /a/, /i/)



# Framework

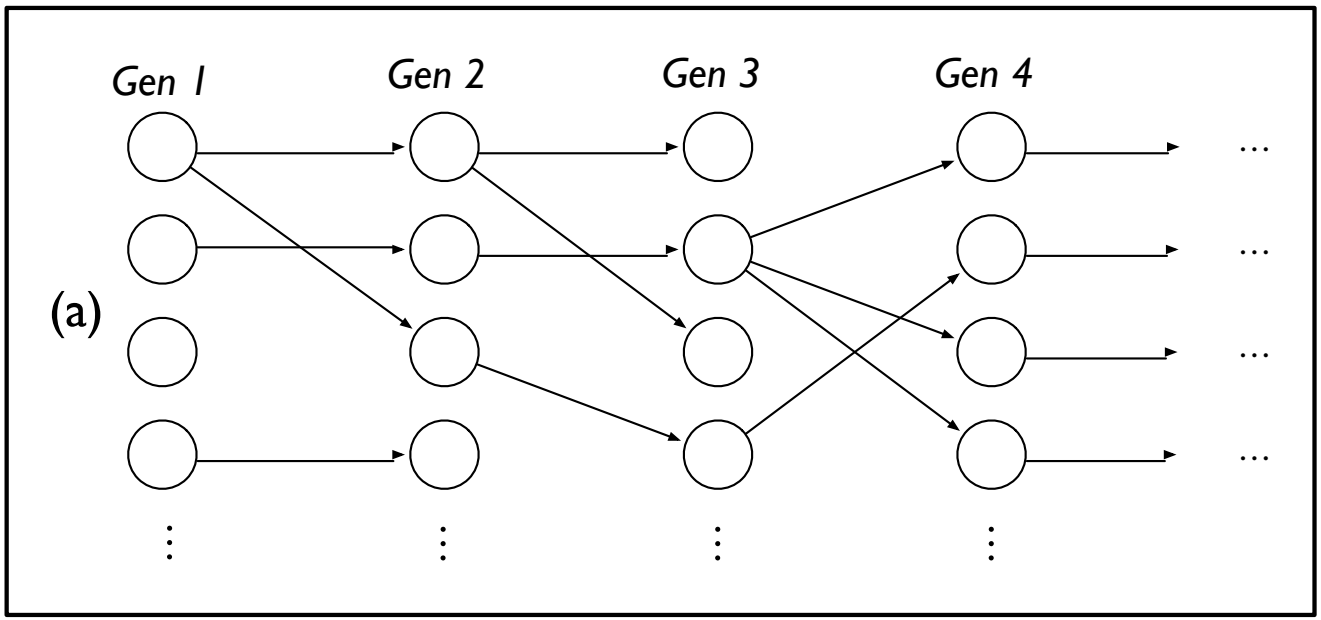
- *Learner's prior:* (strength of) **categoricity bias** (CB)



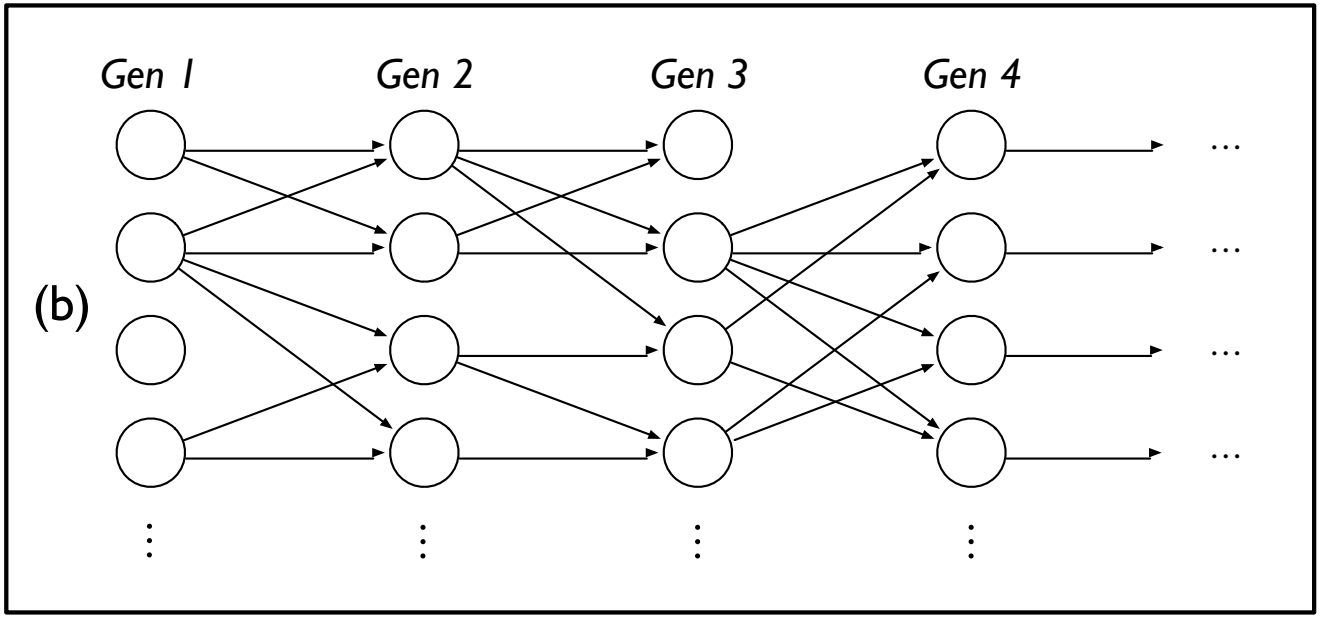
# Framework

- *Population structure*: Single teacher ($m = 1$)



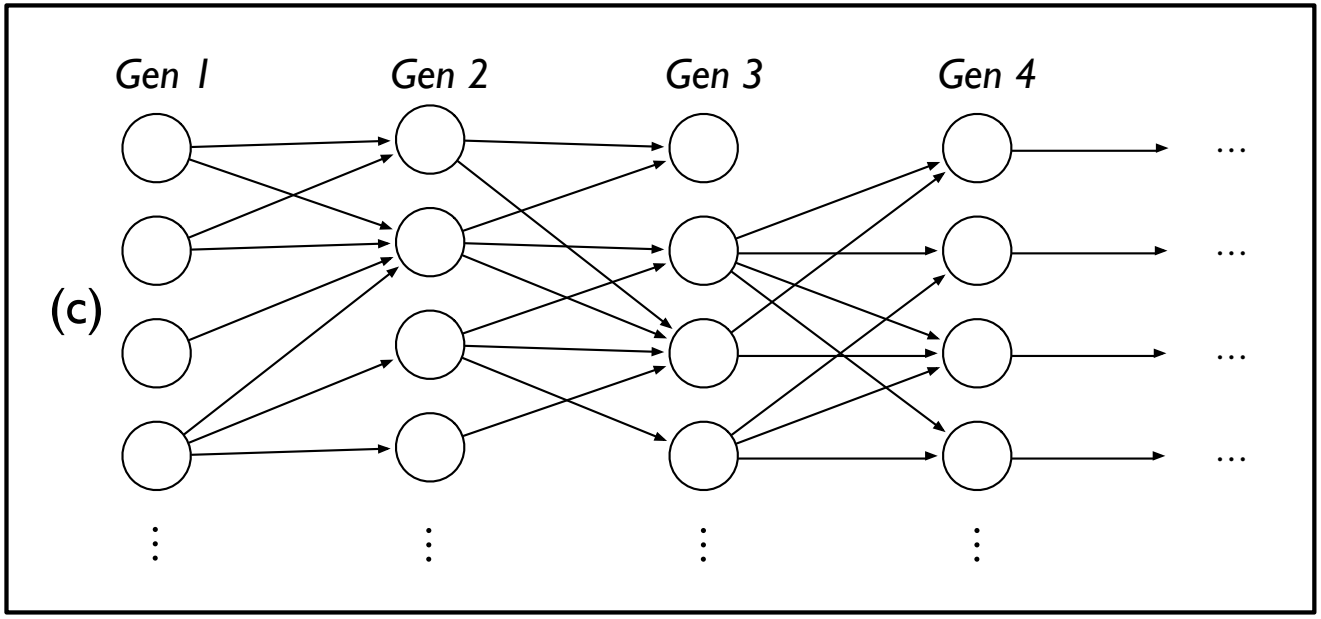
# Framework

- *Population structure*: Multiple teachers ($m = 2$)



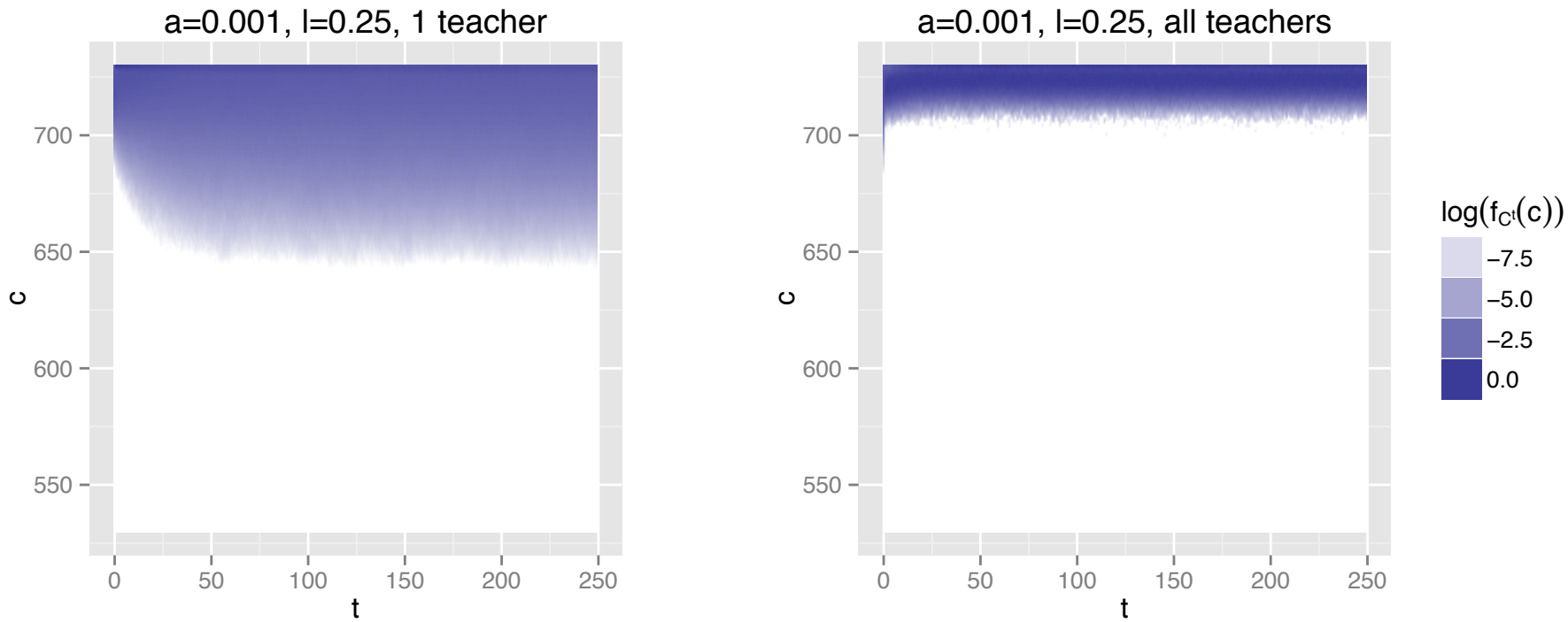
# Framework

- *Population structure*: Multiple teachers ($m = M$)

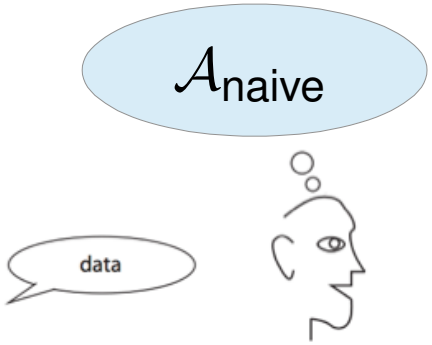


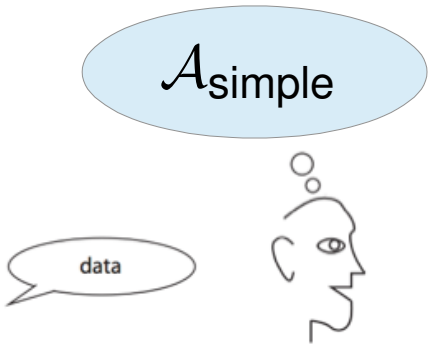
# Framework

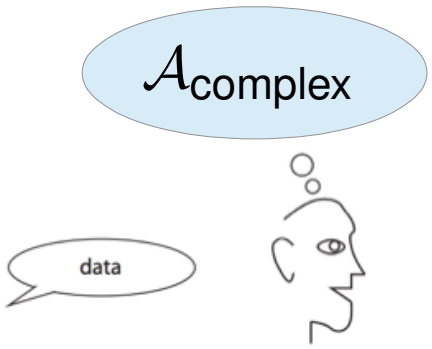
- *Outcome:* distribution of $C$ in the population at time $t$ ($E[C^t]$, $\mathrm{Var}(C^t)$)



# 6 models: $3 \times 2$



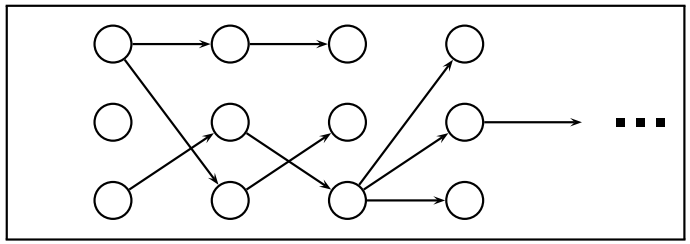




$\longleftarrow$ 3 learning strategies

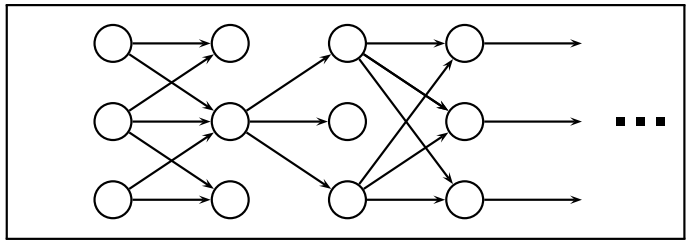
2 kinds of population

$\downarrow$

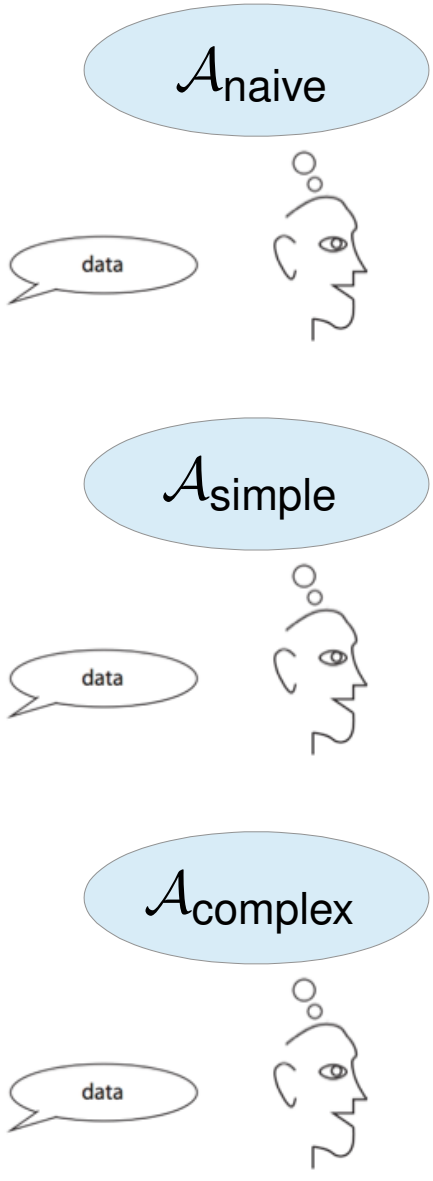
$\mathcal{S}_{\text{single}}$



$\mathcal{S}_{\text{multiple}}$



# 6 models: $3 \times 2$

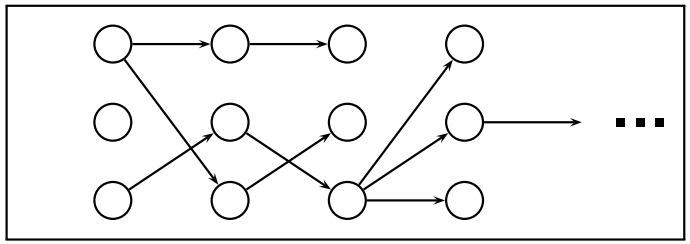



No within-gen learning

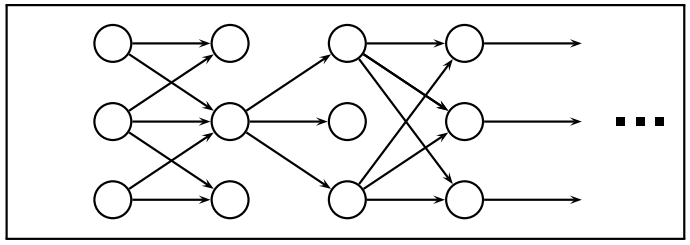
populations are 'large'

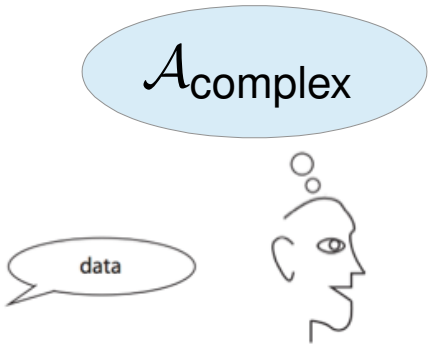
$\downarrow$

$\mathcal{S}_{\text{single}}$



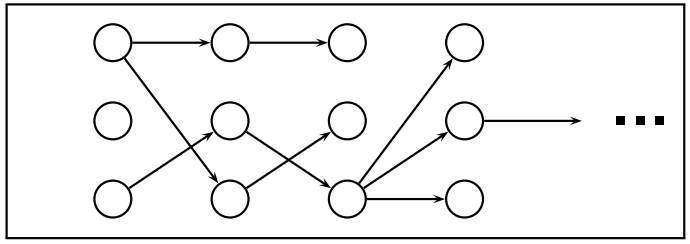

$\mathcal{S}_{\text{multiple}}$



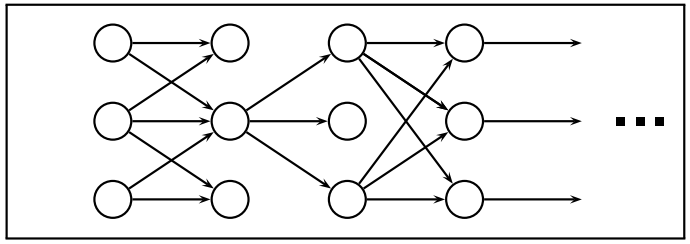



## Today: focus on $\mathcal{A}_{\text{complex}}$

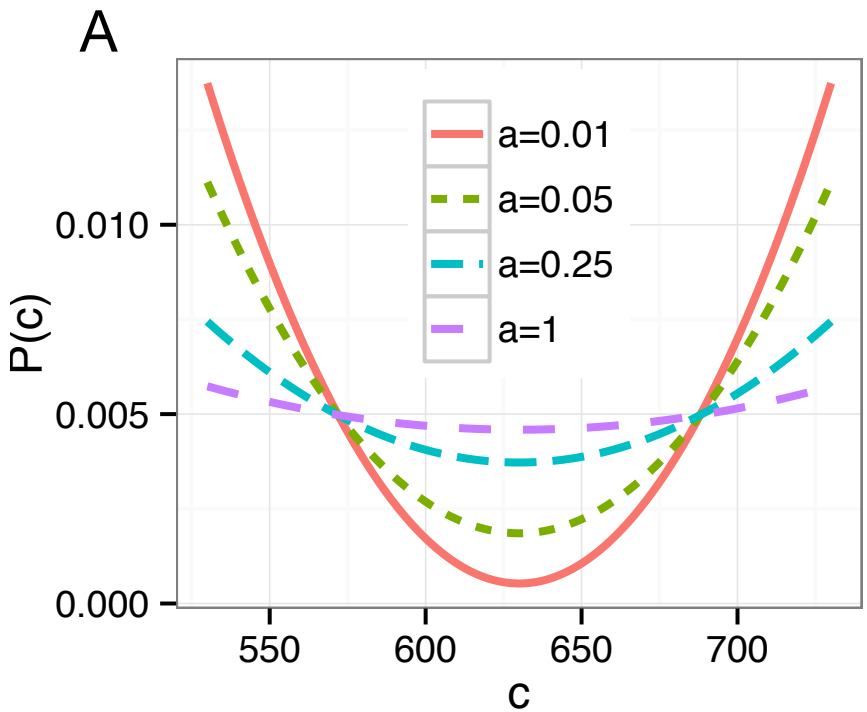
$\mathcal{S}_{\text{single}}$




$\mathcal{S}_{\text{multiple}}$



# $\mathcal{A}_{\text{complex}}$: quadratic polynomial

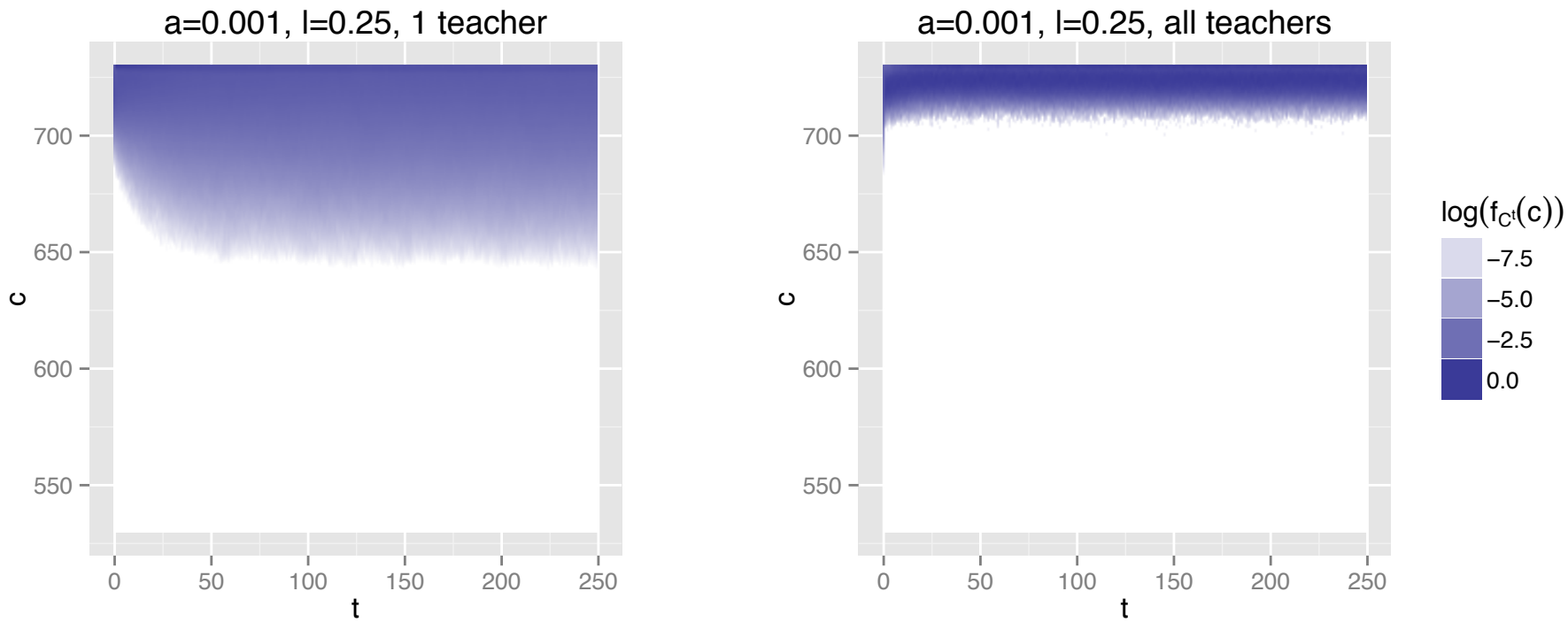



- *a*: 'flatness' of prior
- $\lambda$: mean of bias distribution
- $\omega$: variance of bias distribution

- `nGens`: number of generations
- `nExamples`: # training examples
- `nLearners`: # of learners per gen
- `teachers`: *M* ('single', 'some', 'all')
- `nTeachers`: # of teachers (when `teachers`=='some')

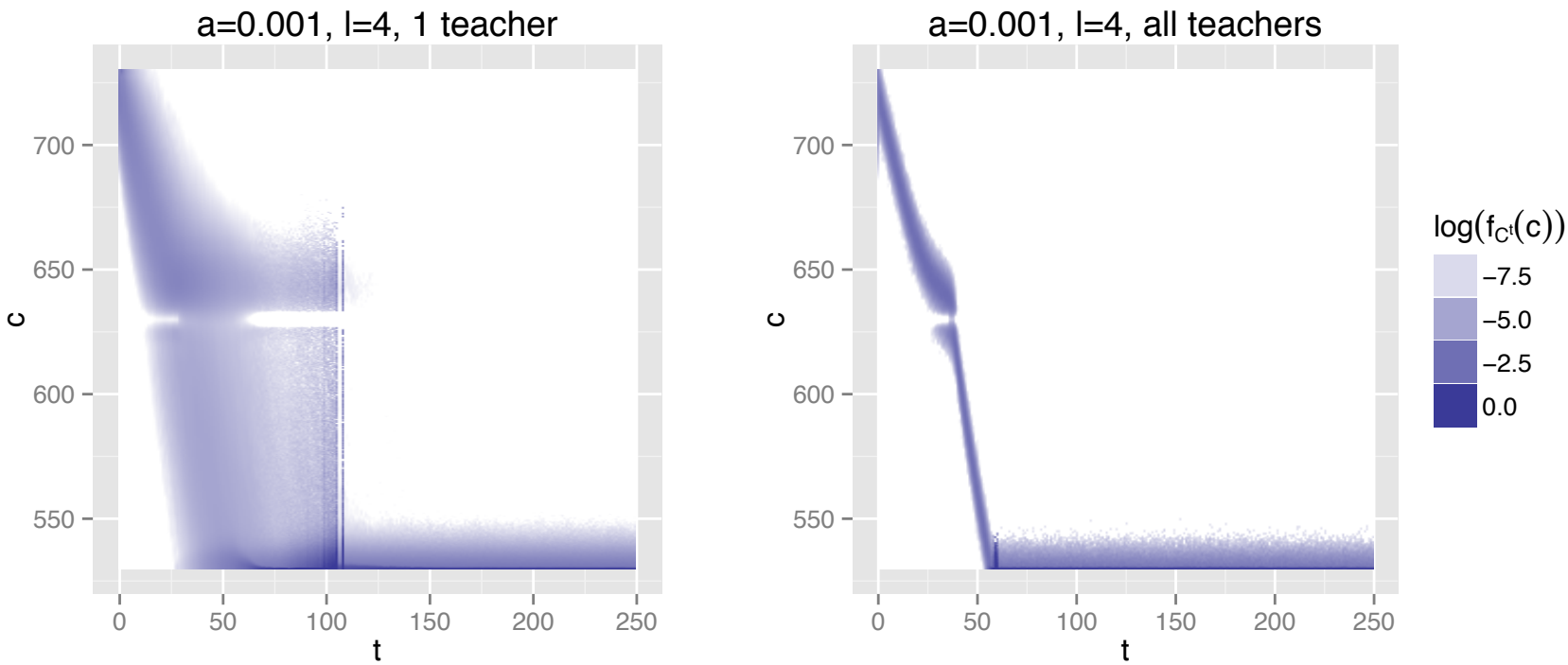
# Evolution of Var[$C^t$]: strong prior, **weak** bias

- For sufficiently strong prior: stable contextual variation



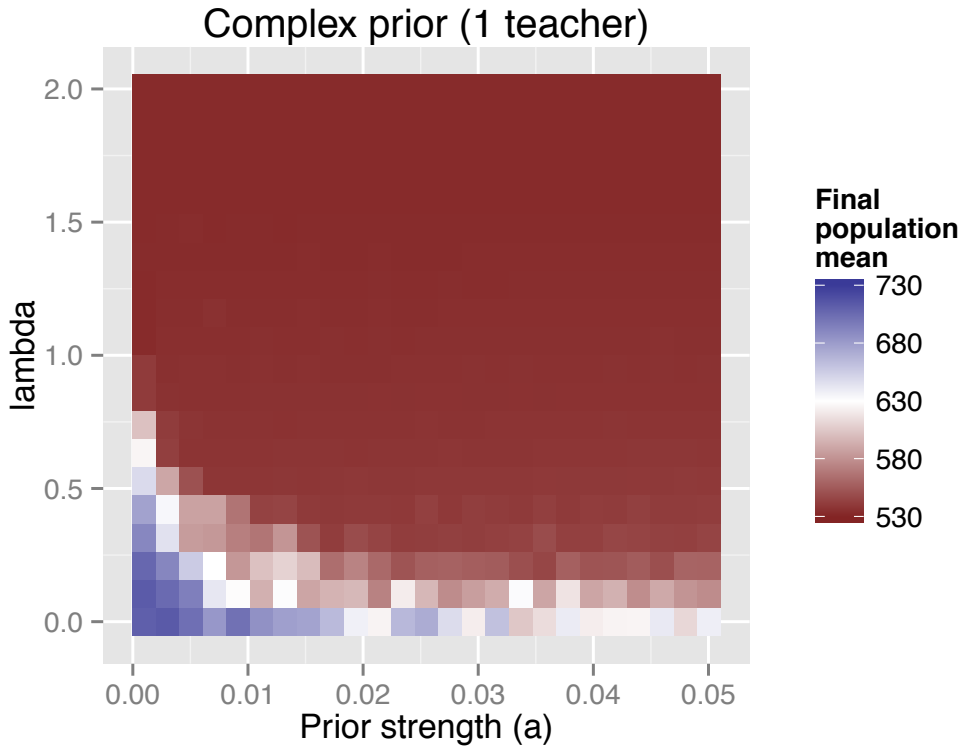
# Evolution of Var[$C^t$]: strong prior, **strong** bias

- For sufficiently strong bias: rapid change to stable umlaut



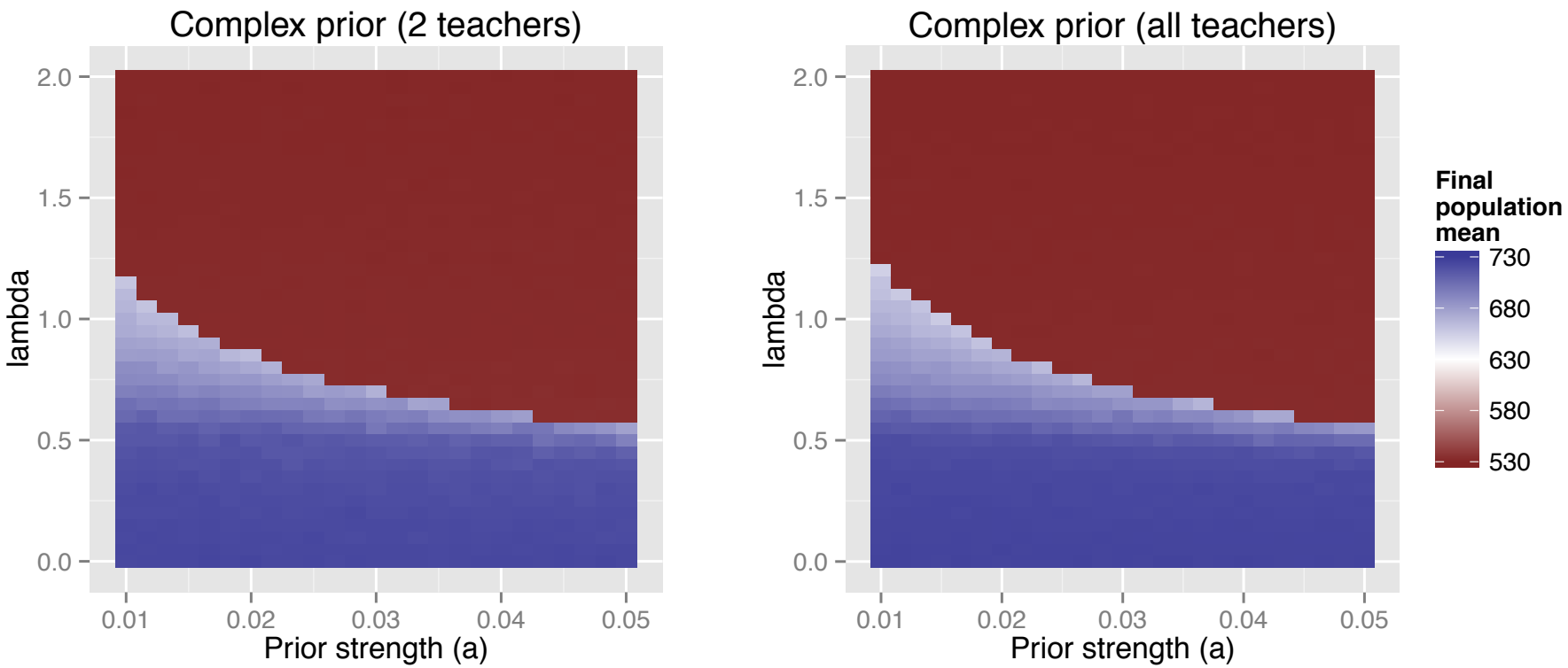
# Evolution of $E[C^t]$, $\mathcal{S}_{\text{single}}$

- Single teacher: stable only when $\lambda = 0$ or very strong prior
- Intermediate distributions of $C^t$ often reflect bimodal distribution (i.e., individuals stable at endpoints)



# Evolution of $E[C^t]$, $\mathcal{S}_{\text{multiple}}$

- Multiple teachers: bifurcation for given prior strength ($a$) once $\lambda$ crosses some critical value
- Required $\lambda \propto a$



# Discussion

- Questions: how do assumptions about transmission bias, categoricity bias, and population structure translate into population-level dynamics of a continuous parameter?

# Discussion

- Questions: how do assumptions about transmission bias, categoricity bias, and population structure translate into population-level dynamics of a continuous parameter?

- First goal: *stability*

# Discussion

- Questions: how do assumptions about transmission bias, categoricity bias, and population structure translate into population-level dynamics of a continuous parameter?
- First goal: *stability*
- Stability is possible even in the presence of bias: transmission bias doesn't entail overapplication
  (cf. WLH, Baker 2008)

# Discussion

- Second goal: *change*, or rather: how does stability give way to change?

# Discussion

- Second goal: *change*, or rather: how does stability give way to change?
  - Bifurcations only observed with complex prior enforcing categoricity (cf. Komarova et al., Nigoyi, etc.)
  - Population structure important in that $m = 1 \neq m > 1$ (cf. Griffiths, S. Kirby, et al.)

# Discussion

- Second goal: *change*, or rather: how does stability give way to change?
  - Bifurcations only observed with complex prior enforcing categoricity (cf. Komarova et al., Nigoyi, etc.)
  - Population structure important in that $m = 1 \neq m > 1$ (cf. Griffiths, S. Kirby, et al.)
- **Q:** for what ranges of parameter settings are these conclusions valid?

# Extensions

- Production bias is the external force most commonly invoked in models of sound change

# Extensions

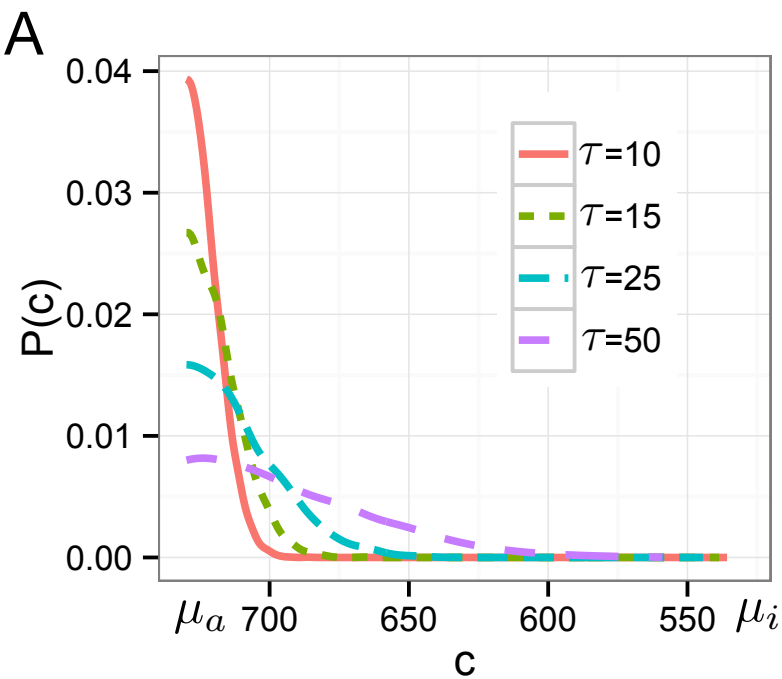
- Production bias is the external force most commonly invoked in models of sound change
- ... but clearly not behind **all** changes: many other factors invoked by (socio)phon(eticians), e.g.
  - Contact (between subpopulations)
  - Social weight (of variants, speakers, groups)
  - Interaction (convergence, divergence)

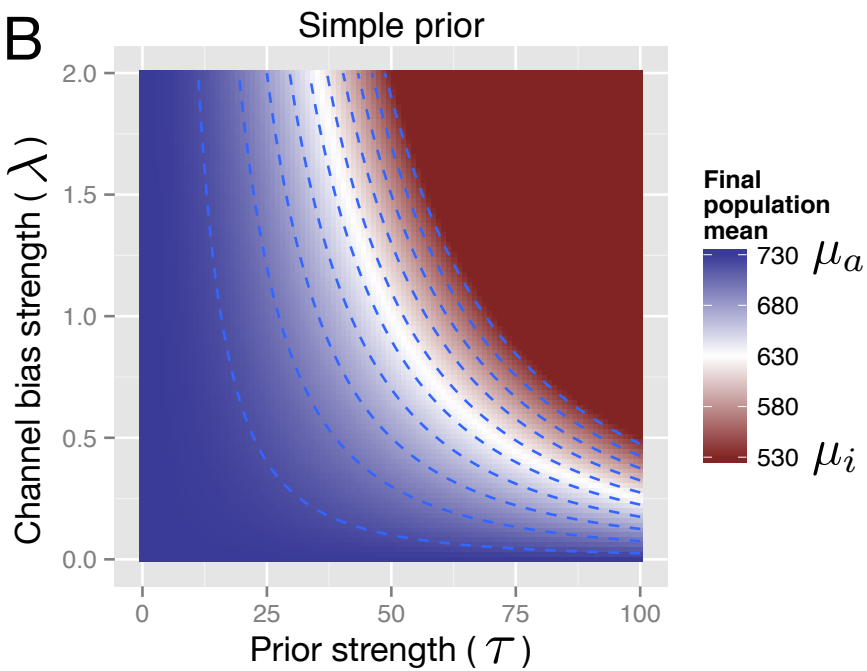
# Extensions

- Production bias is the external force most commonly invoked in models of sound change
- ... but clearly not behind **all** changes: many other factors invoked by (socio)phon(eticians), e.g.
  - Contact (between subpopulations)
  - Social weight (of variants, speakers, groups)
  - Interaction (convergence, divergence)
- Next set of questions:
  1. Does using production bias as the external force have a unique dynamics?
  2. If not, will *any* kind of external force produce the same behaviour at the population level?

# $\mathcal{A}_{\text{simple}} : \tau \sim \mathcal{N}(0, \tau^2)$

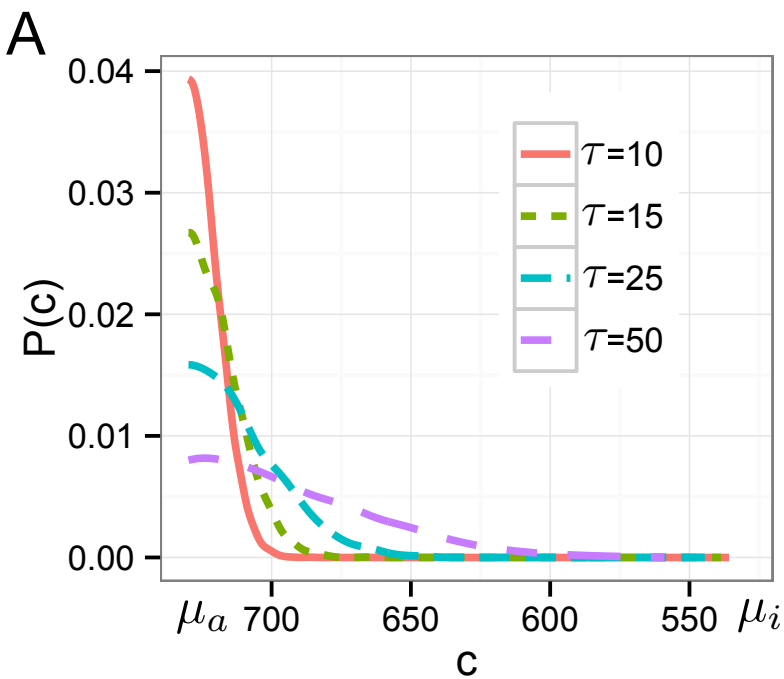
- Assume learner has prior with variance $\tau$, takes MAP estimate of $\hat{c}$
- $\mu$ *always* moves to stable value, regardless of population structure (*m*)

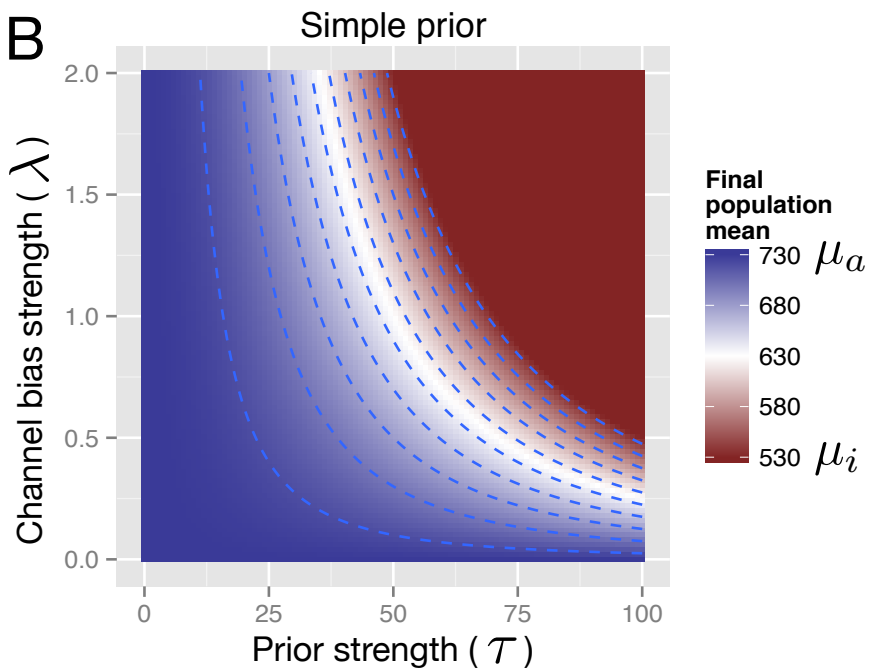


# $\mathcal{A}_{\text{simple}} : \tau \sim \mathcal{N}(0, \tau^2)$

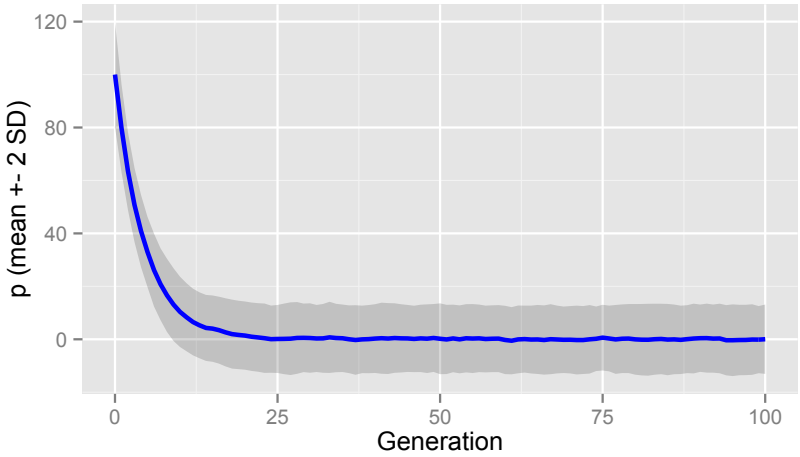
- Assume learner has prior with variance $\tau$, takes MAP estimate of $\hat{c}$
- Variance rapidly stabilizes (although final value depends on $m$)



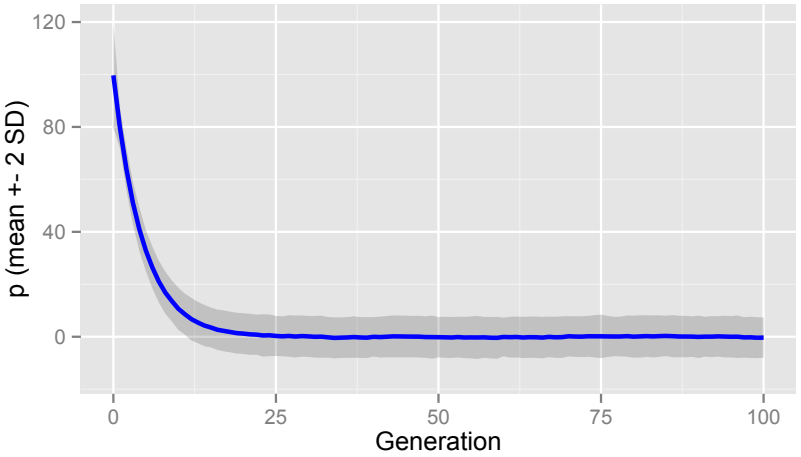



# Evolution of $\pi_t$ under $\mathcal{A}_{\text{simple}}$, $c_0 \sim N, \tau = 10$

$\mathcal{S}_{\text{single}}$



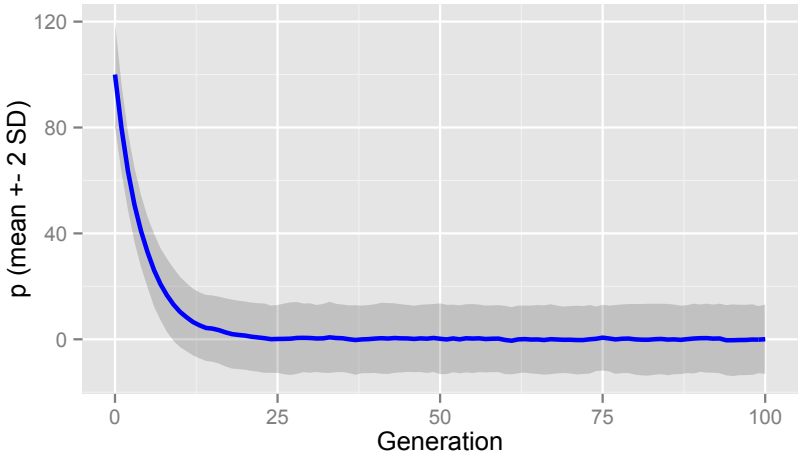

$\mathcal{S}_{\text{multiple}}$



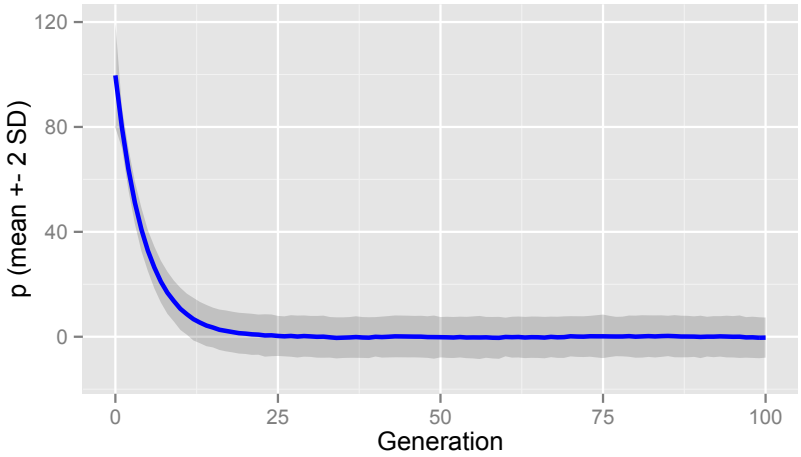

- $\mathcal{A}_{\text{simple}}$ can't model any change (Goal 3)
- Pop. structure impacts rate but not qualitative outcome

# Evolution of $\pi_t$ under $\mathcal{A}_{\text{simple}}, c_0 \sim N, \tau = 10$

$\mathcal{S}_{\text{single}}$




$\mathcal{S}_{\text{multiple}}$

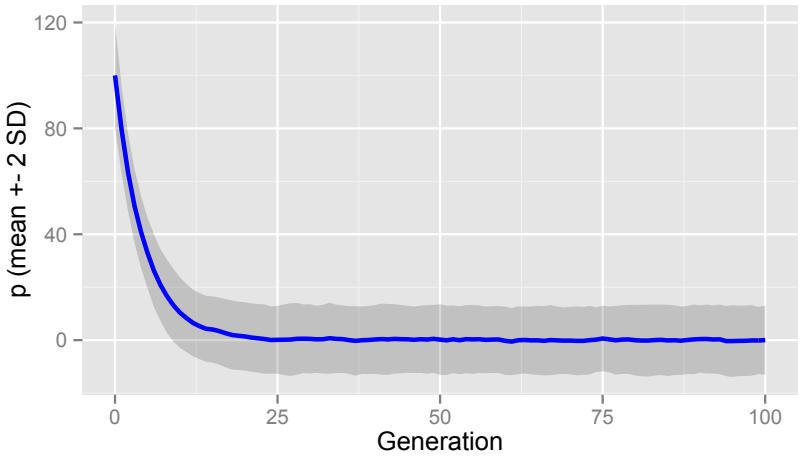



- $\mathcal{A}_{\text{simple}}$ can't model any change (Goal 3)
- Pop. structure impacts rate but not qualitative outcome

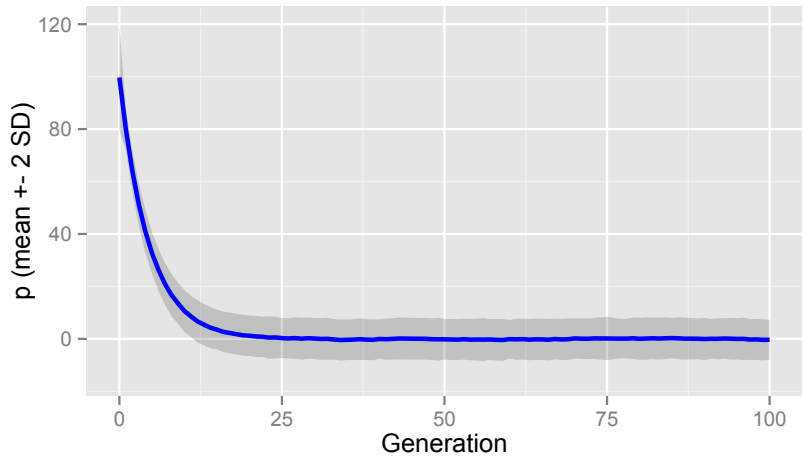
No threshold in system params gives change to stable umlaut

# Evolution of $\pi_t$ under $\mathcal{A}_{\text{simple}}$, $c_0 \sim N, \tau = 10$

$\mathcal{S}_{\text{single}}$




$\mathcal{S}_{\text{multiple}}$




- $\mathcal{A}_{\text{simple}}$ can't model any change (Goal 3)
- Population structure impacts rate but not outcome

Prior needs to encode some kind of category preference.