Universität Zürich[UZH]
Department of General Linguistics

# Areal diachronies

Balthasar Bickel

# Basic assumptions

## Basic assumptions

All area effects are effects on diachrony:

# Basic assumptions

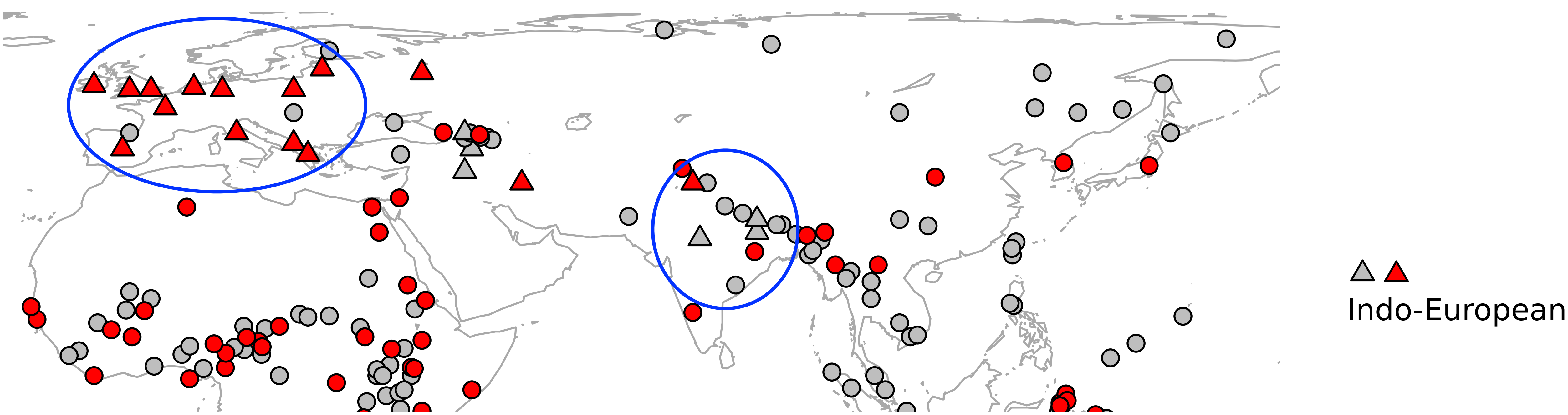
All area effects are effects on diachrony:

- they take place over time $\rightarrow$ need methods for estimating the diachronic process that led to areas

All area effects are effects on diachrony:

- they take place over time → need methods for estimating the diachronic process that led to areas
- they can involve innovation ***and*** retention alike

for example, gender systems tend to cluster areally not by innovation but by retention (Nichols 2003): pronominal gender (Siewierska 2005)



# Methodological challenge

# Methodological challenge

How to detect areal signals of **retention and innovation**?

# Methodological challenge

How to detect areal signals of **retention and innovation**?

- Since retention and innovation rates are known to vary *extremely* across time and space,

  *let's not try to estimate them!*

## Methodological challenge

How to detect areal signals of **retention and innovation**?

- Since retention and innovation rates are known to vary *extremely* across time and space,

  *let's not try to estimate them!*

- Since we know virtually no proto-language, methods based on reconstructions and/or tree structures are difficult to evaluate and calibrate, so

  *let's not try to estimate the structures of proto-languages and/or trees!*

How to detect areal signals of **retention and innovation**?

- Since retention and innovation rates are known to vary *extremely* across time and space,

  *let's not try to estimate them!*

- Since we know virtually no proto-language, methods based on reconstructions and/or tree structures are difficult to evaluate and calibrate, so

  *let's not try to estimate the structures of proto-languages and/or trees!*

- But let's estimate diachronic trends, nevertheless!

How to detect areal signals of **retention and innovation**?

- Since retention and innovation rates are known to vary *extremely* across time and space,

  *let's not try to estimate them!*

- Since we know virtually no proto-language, methods based on reconstructions and/or tree structures are difficult to evaluate and calibrate, so

  *let's not try to estimate the structures of proto-languages and/or trees!*

- But let's estimate diachronic trends, nevertheless!
- And do so without neglecting isolates and small families!

# The Family Bias Method

# The Family Bias Method

- Solution comes from the observation that we actually don't need rate estimates and proto-languages at all:

## The Family Bias Method

- Solution comes from the observation that we actually don't need rate estimates and proto-languages at all:
  - For picking up area signals, the difference between retention and innovation does not matter:

# The Family Bias Method

- Solution comes from the observation that we actually don't need rate estimates and proto-languages at all:
  - For picking up area signals, the difference between retention and innovation does not matter:
    - languages may prefer to keep X more inside than outside an area

# The Family Bias Method

- Solution comes from the observation that we actually don't need rate estimates and proto-languages at all:
  - For picking up area signals, the difference between retention and innovation does not matter:
    - languages may prefer to keep X more inside than outside an area
    - languages may prefer to innovate X more inside than outside an area

# The Family Bias Method

- Solution comes from the observation that we actually don't need rate estimates and proto-languages at all:
  - For picking up area signals, the difference between retention and innovation does not matter:
    - languages may prefer to keep X more inside than outside an area
    - languages may prefer to innovate X more inside than outside an area
  - The synchronic result is the same: we have bias towards X in the end.

# The Family Bias Method

Synchronic observations
on *demonstrably related*
languages:

| | | | |
|---|---|---|---|
| X | X | X | X |
| X | X | X | X |
| Y | | | |

# The Family Bias Method

Synchronic observations on *demonstrably related* languages:

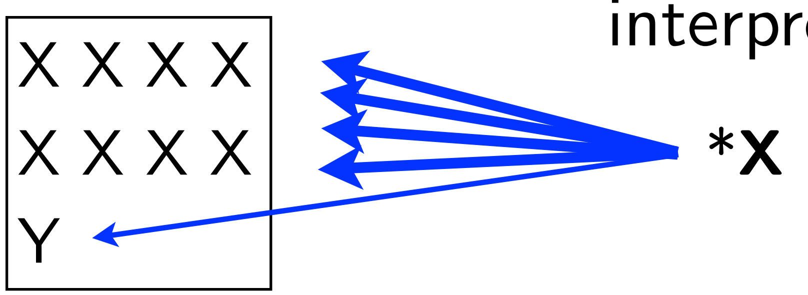
| | | | |
|---|---|---|---|
| X | X | X | X |
| X | X | X | X |
| Y | | | |

Possible diachronic interpretations:

# The Family Bias Method

Synchronic observations on *demonstrably related* languages:

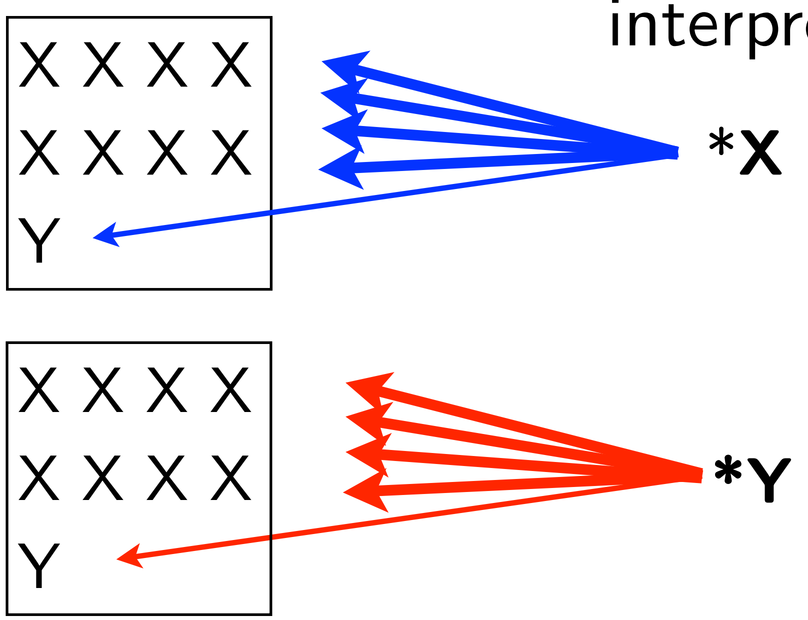
Possible diachronic interpretations:



# The Family Bias Method

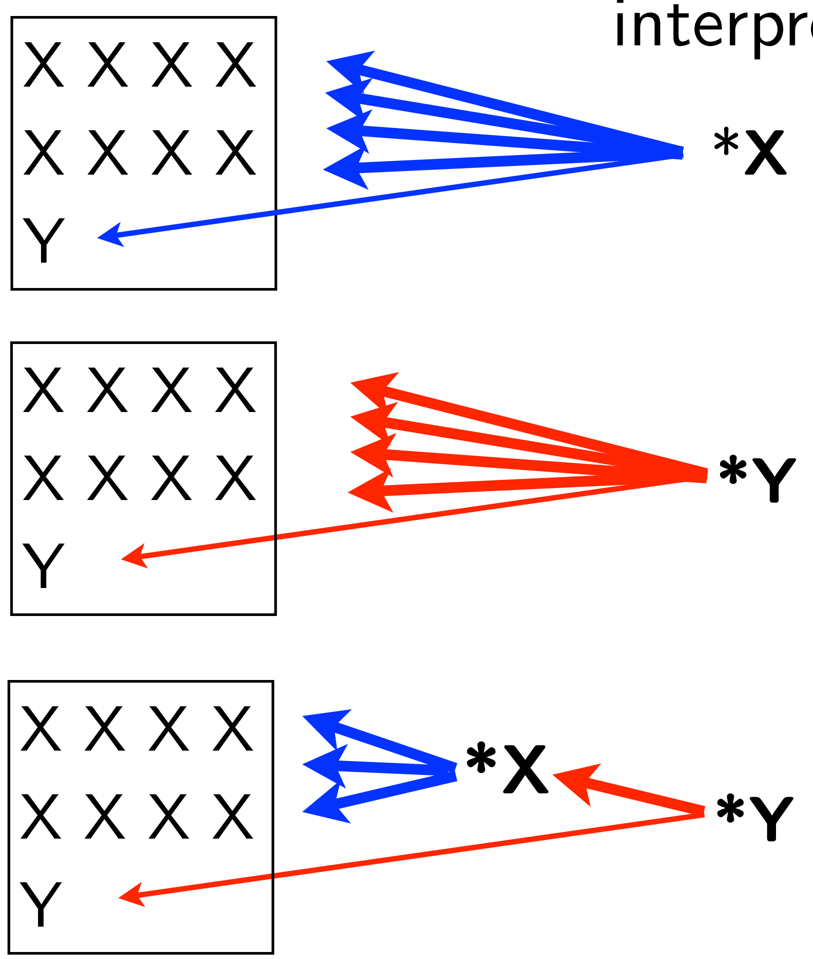
Synchronic observations on *demonstrably related* languages:

Possible diachronic interpretations:



Synchronic observations on *demonstrably related* languages:

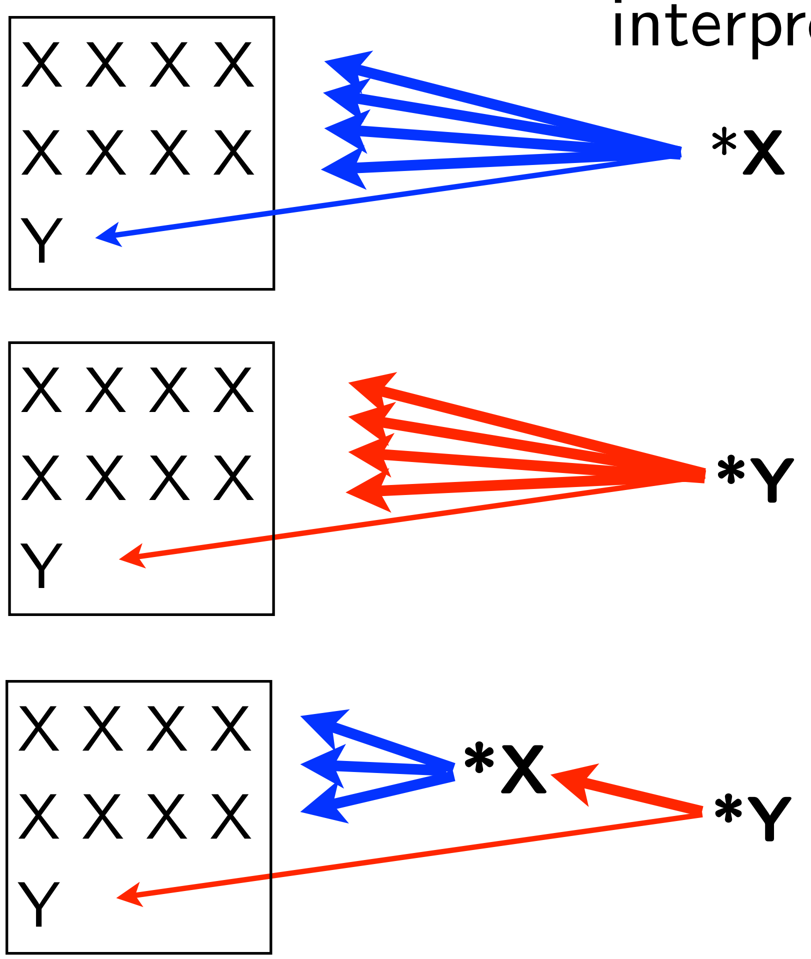
Possible diachronic interpretations:



Synchronic observations on *demonstrably related* languages:

Possible diachronic interpretations:

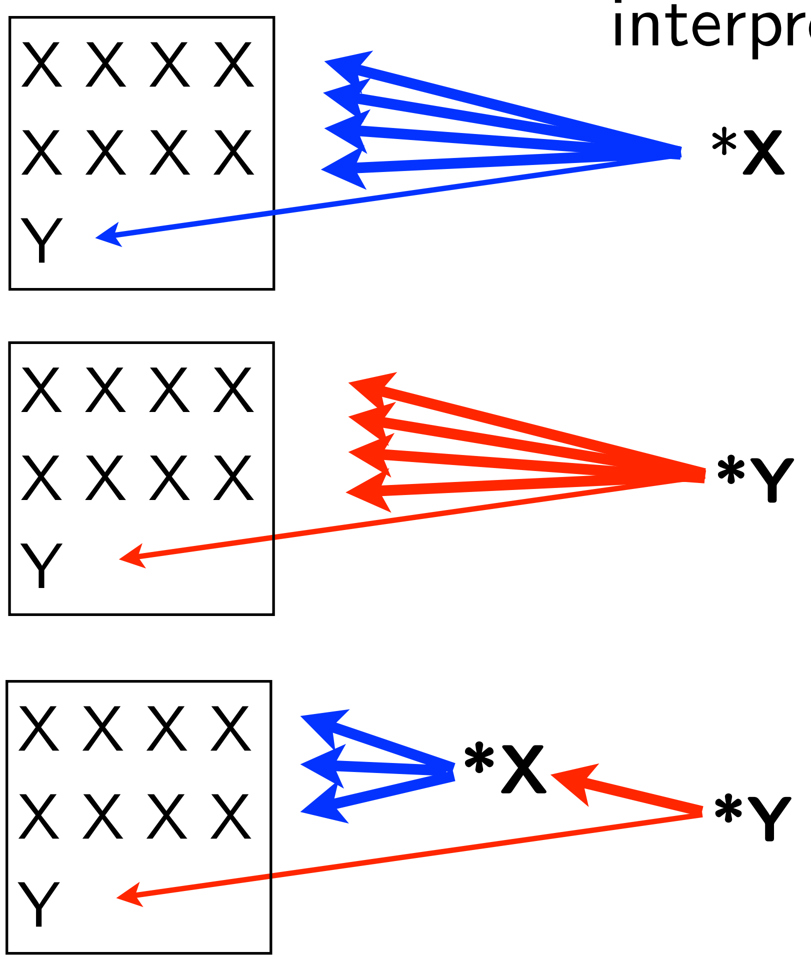
Conclusion: different probabilities of innovation *and* retention



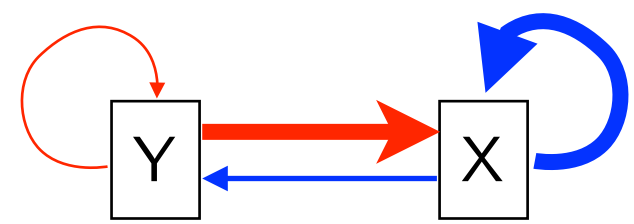
# The Family Bias Method

Synchronic observations on *demonstrably related* languages:

Possible diachronic interpretations:

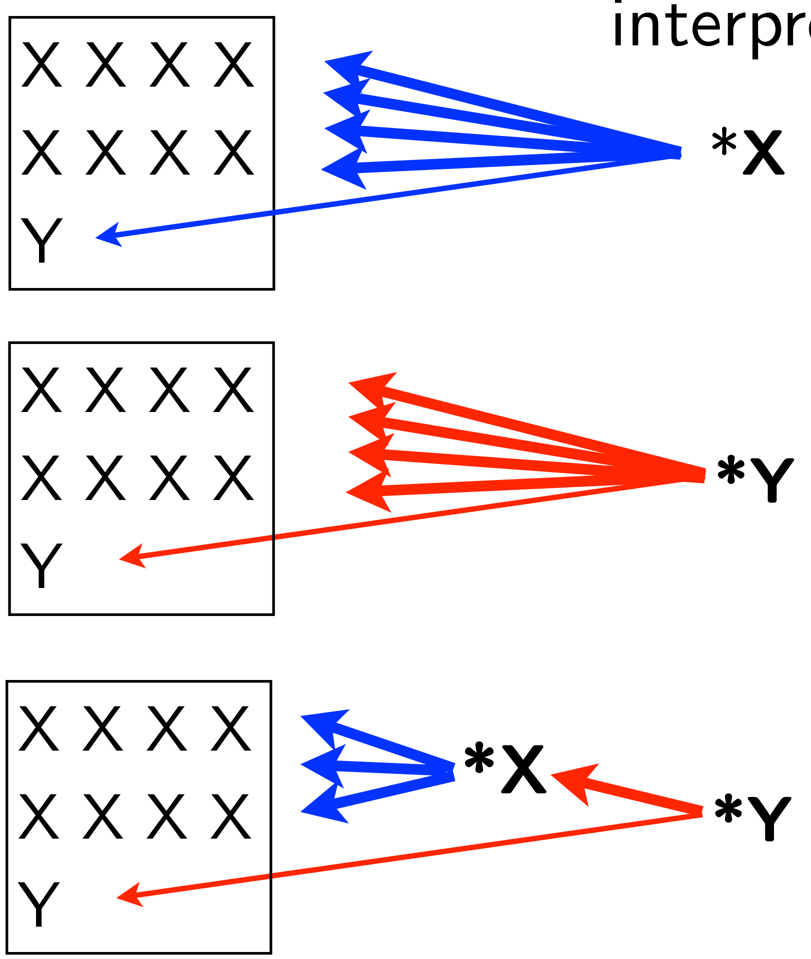



Conclusion: different probabilities of innovation *and* retention

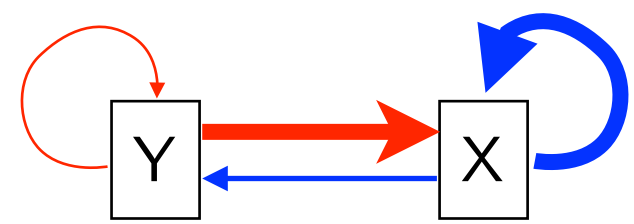


Synchronic observations on *demonstrably related* languages:

Possible diachronic interpretations:



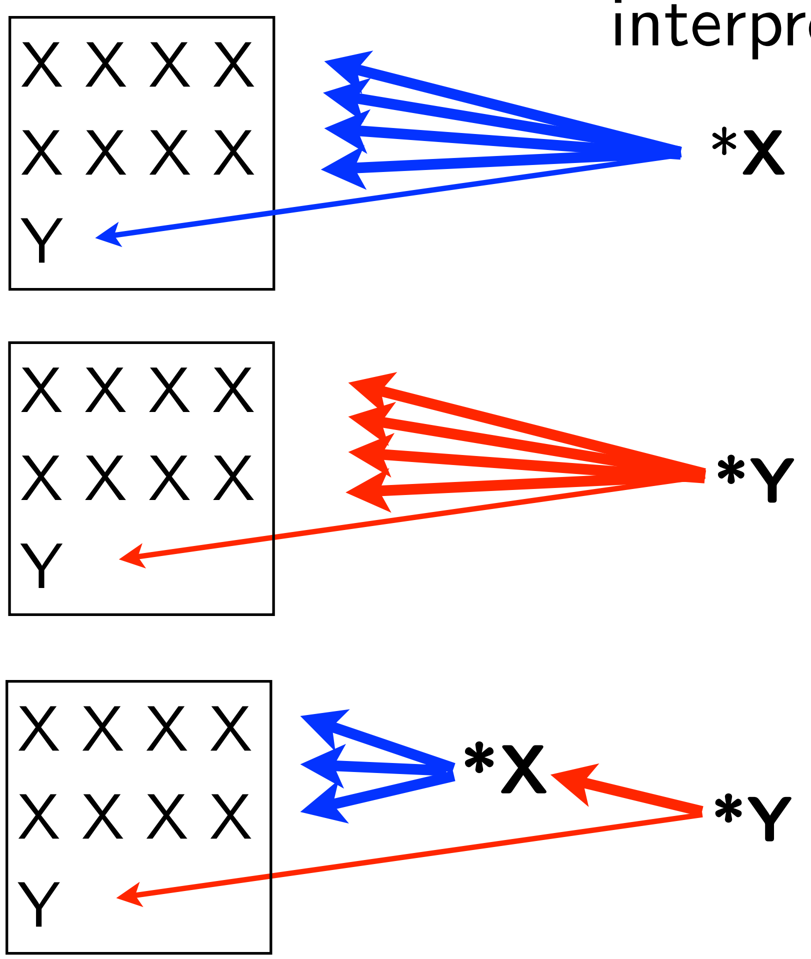

Conclusion: different probabilities of innovation *and* retention



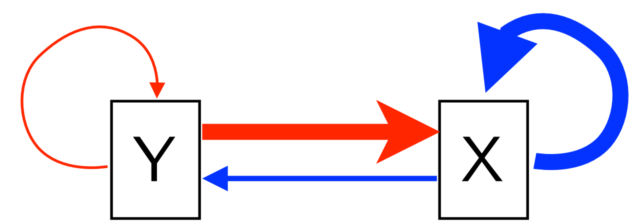

$$Pr(Y>X) > Pr(X>Y)$$

Synchronic observations on *demonstrably related* languages:

Possible diachronic interpretations:




Conclusion: different probabilities of innovation *and* retention



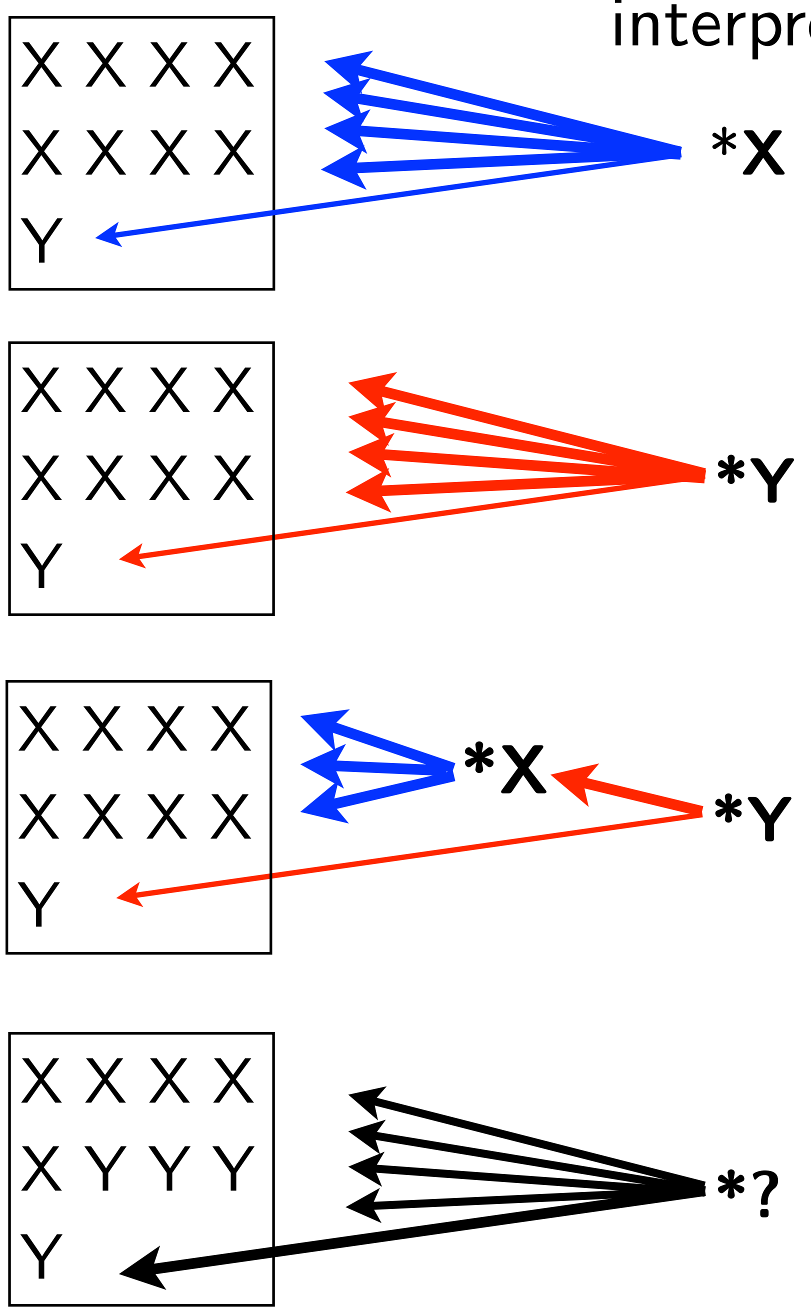

$$Pr(Y{>}X) > Pr(X{>}Y)$$

("**Family Bias**")

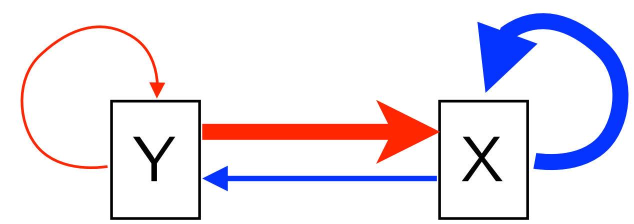
# The Family Bias Method

Synchronic observations on *demonstrably related* languages:

Possible diachronic interpretations:




Conclusion: different probabilities of innovation *and* retention



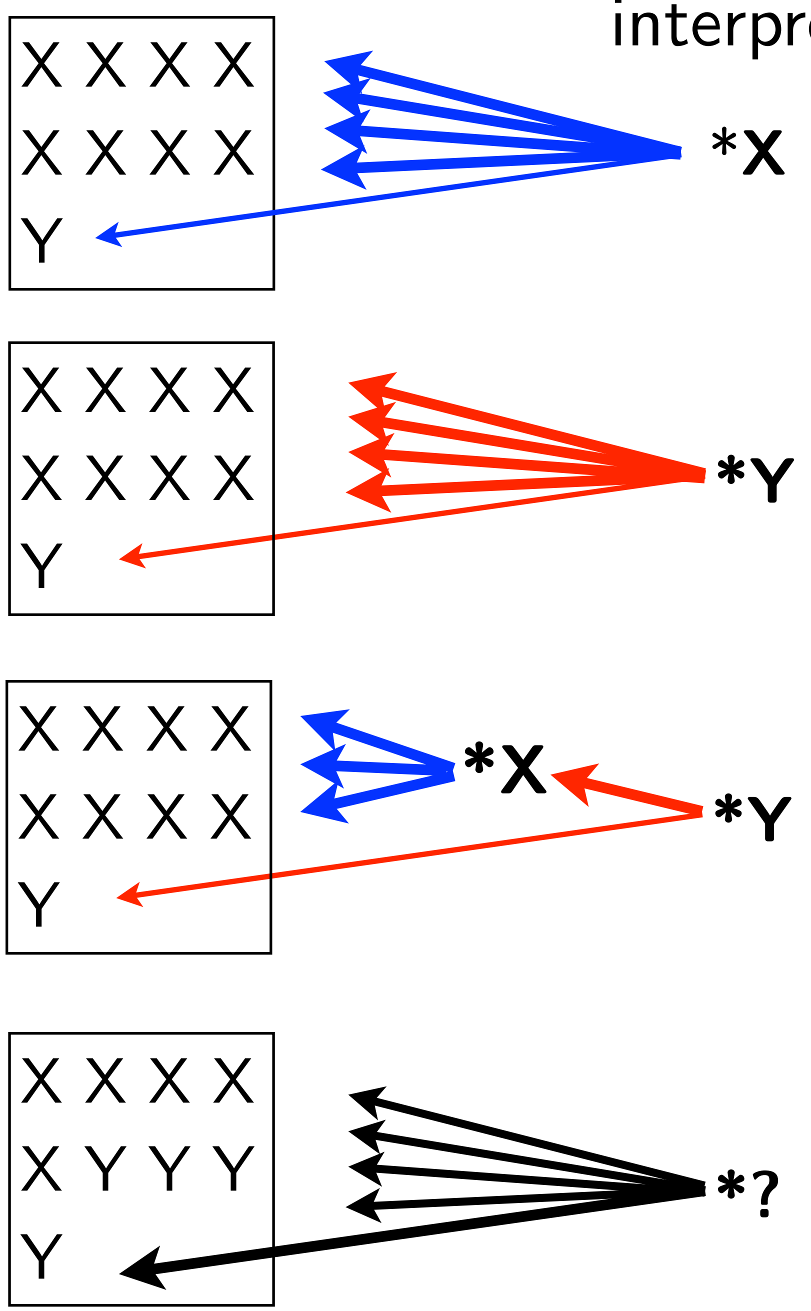

$Pr(Y{>}X) > Pr(X{>}Y)$

("**Family Bias**")

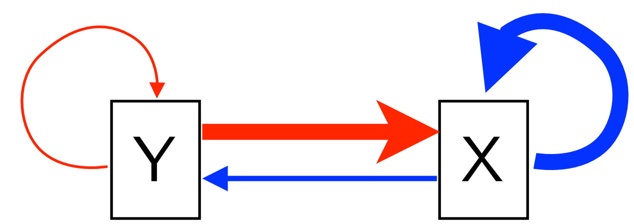
# The Family Bias Method

Synchronic observations on *demonstrably related* languages:

Possible diachronic interpretations:




Conclusion: different probabilities of innovation *and* retention




$Pr(Y>X) > Pr(X>Y)$

(“**Family Bias**”)

$Pr(Y>X) \approx Pr(X>Y)$

(“no bias”, “diverse”)

# Justification of the key assumption of the Family Bias Method

- Only alternative interpretations of a synchronic bias:

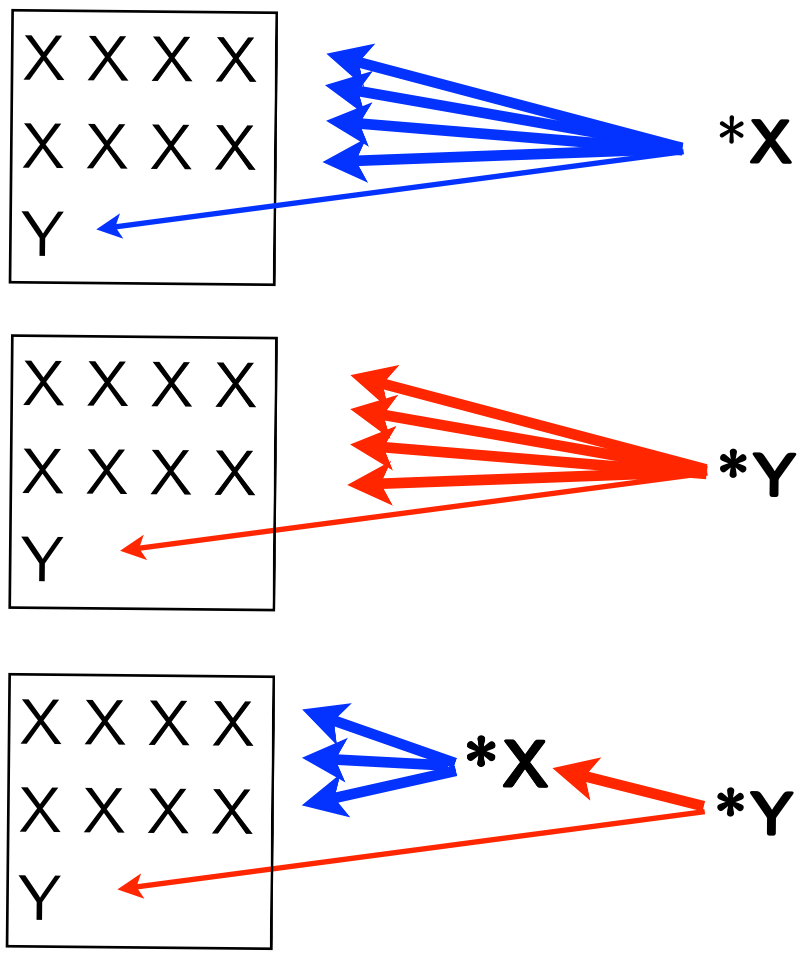
## Justification of the key assumption of the Family Bias Method

- Only alternative interpretations of a synchronic bias:
  (a) deny it: a significant preference for X doesn't tell us anything

# Justification of the key assumption of the Family Bias Method

- Only alternative interpretations of a synchronic bias:
  (a) deny it: a significant preference for X doesn't tell us anything
  (b) assume extreme stability

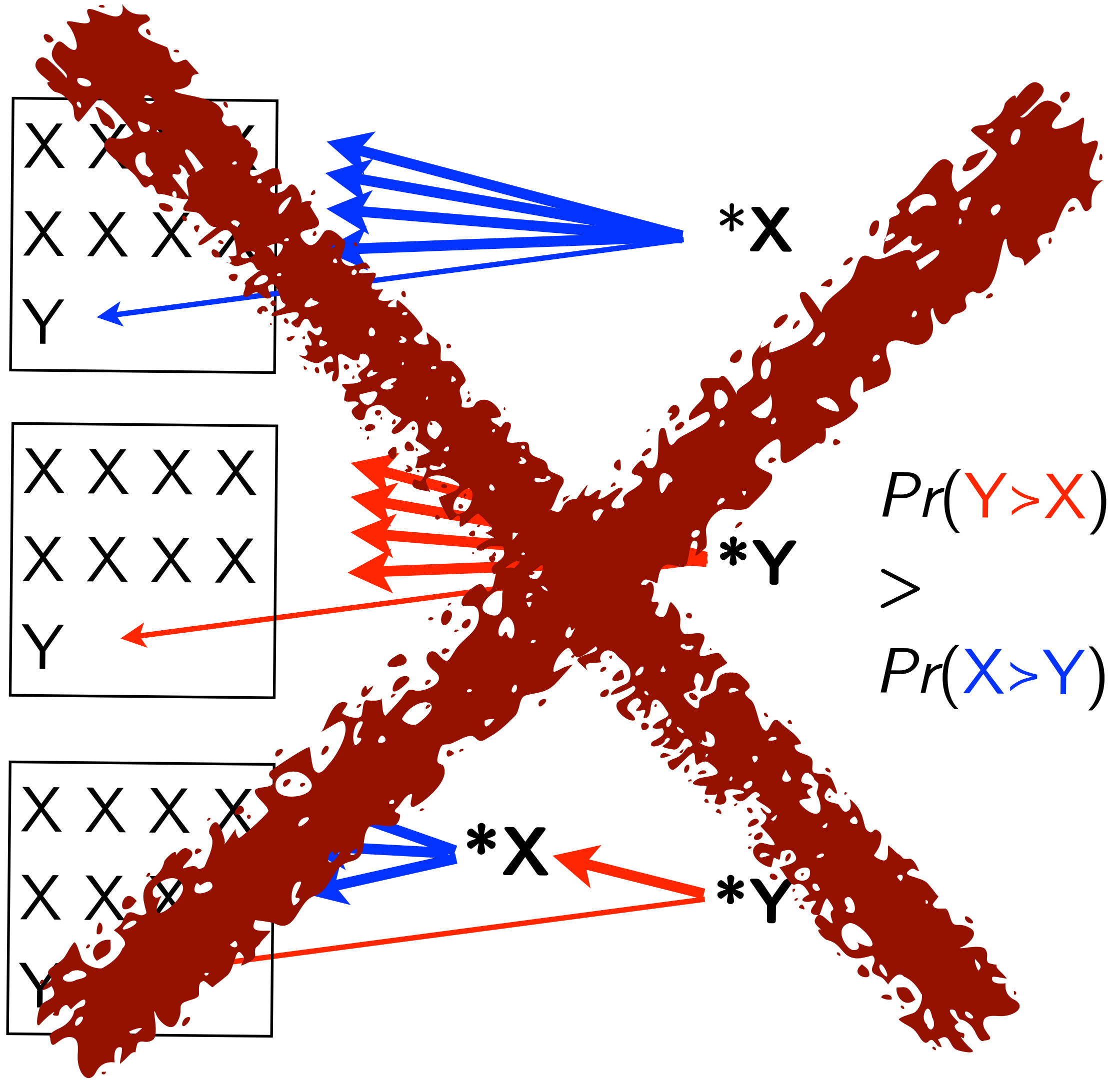
- Only alternative interpretations of a synchronic bias:
  - (a) deny it: a significant preference for X doesn't tell us anything
  - (b) assume extreme stability



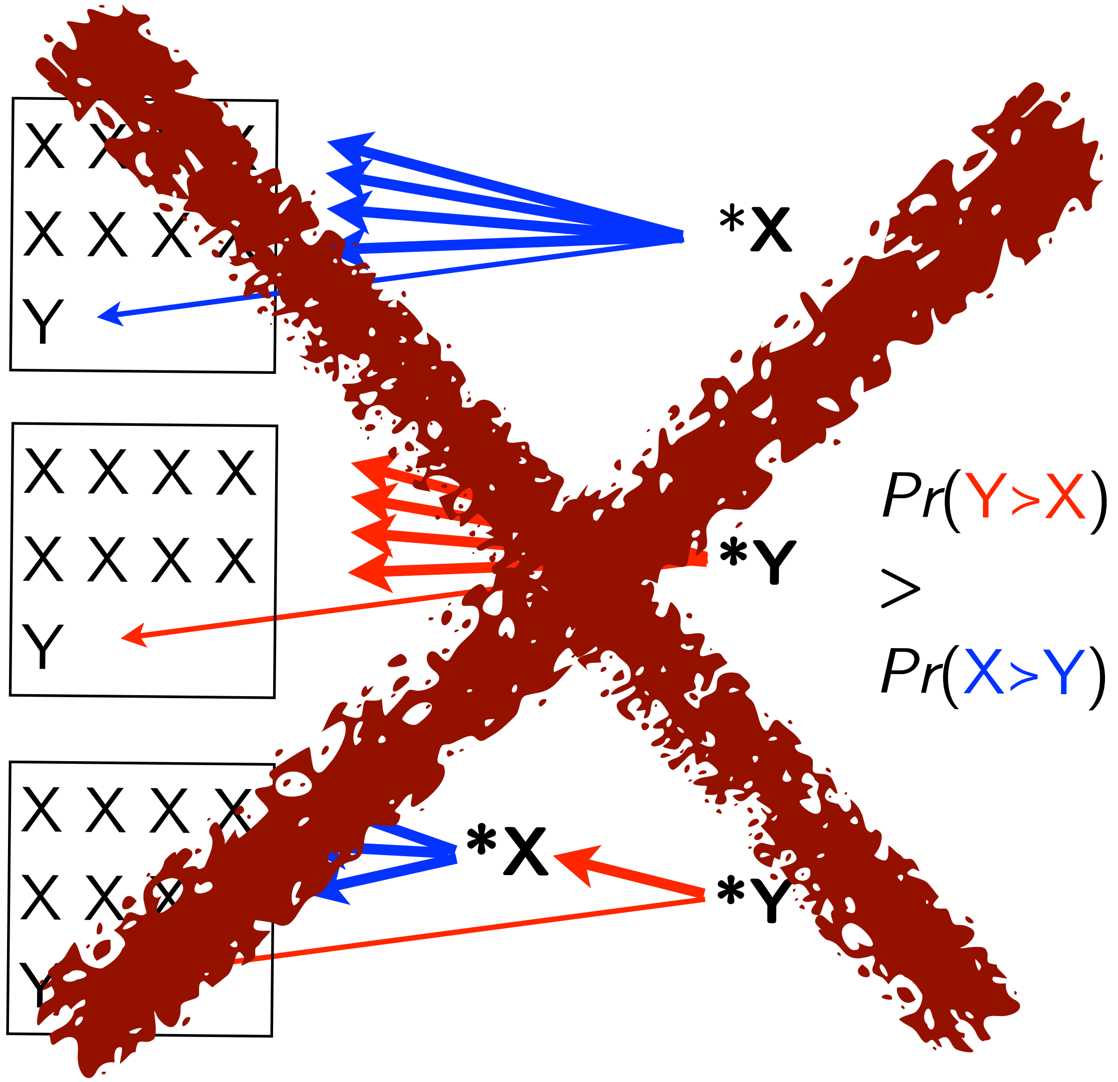

*Pr*(Y>X)
>
*Pr*(X>Y)

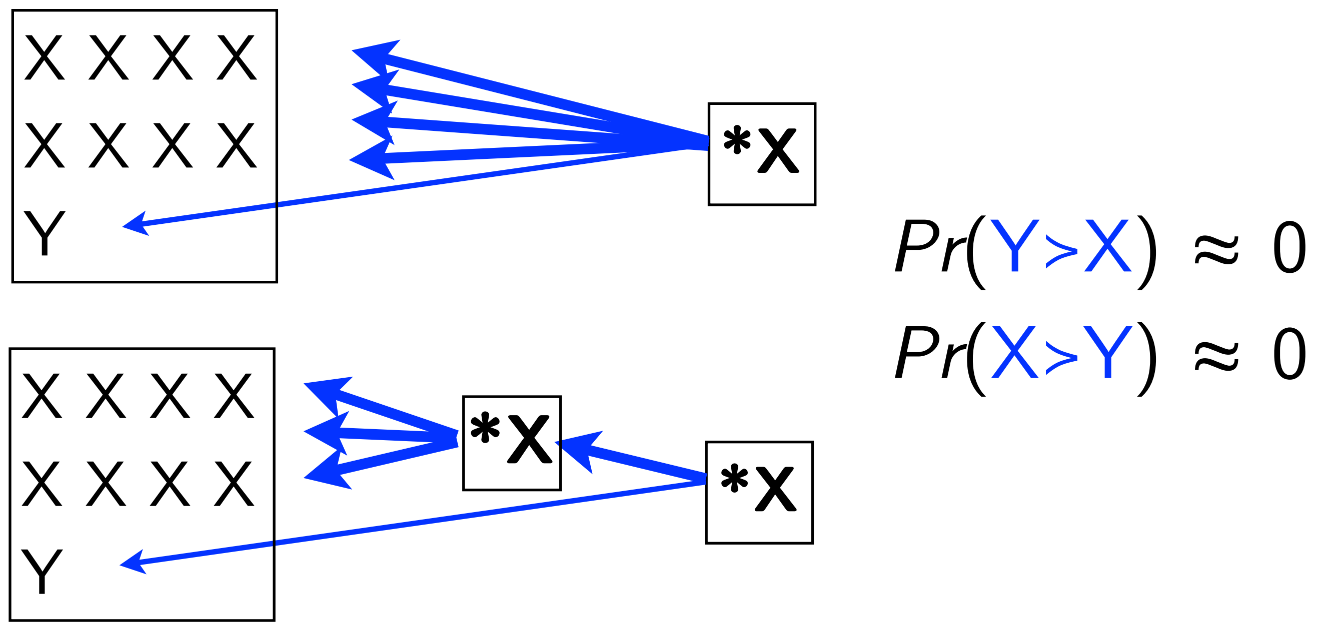
## Justification of the key assumption of the Family Bias Method

- Only alternative interpretations of a synchronic bias:
  - (a) deny it: a significant preference for X doesn't tell us anything
  - (b) assume extreme stability



- Only alternative interpretations of a synchronic bias:
  - (a) deny it: a significant preference for X doesn't tell us anything
  - (b) assume extreme stability



# Justification of the key assumption of the Family Bias Method

## Justification of the key assumption of the Family Bias Method

- In fact, the assumption of extreme stability is behind

- In fact, the assumption of extreme stability is behind
  - the traditional call for "genealogically balanced sampling" (e.g. Dryer 1989) and also

# Justification of the key assumption of the Family Bias Method

- In fact, the assumption of extreme stability is behind
  - the traditional call for "genealogically balanced sampling" (e.g. Dryer 1989) and also
  - all attempts to "control for genealogical relatedness" by building families into statistical models as control factors (e.g. Bickel et al. 2008, Jaeger et al. 2011)

## Justification of the key assumption of the Family Bias Method

- In fact, the assumption of extreme stability is behind
  - the traditional call for "genealogically balanced sampling" (e.g. Dryer 1989) and also
  - all attempts to "control for genealogical relatedness" by building families into statistical models as control factors (e.g. Bickel et al. 2008, Jaeger et al. 2011)
- But typological variables are not remotely as stable as would be required for this ...

# Is $Pr$(Y>X) ≈ $Pr$(X>Y) ≈ 0 (extreme stability) plausible?

# Is $Pr(Y>X) \approx Pr(X>Y) \approx 0$ (extreme stability) plausible?

- $\mathrm{Pr}(Y>X) \approx \mathrm{Pr}(X>Y) \approx 0$ means that changes are extremely unlikely within short time intervals such as those of known families

# Is *Pr*(Y>X) ≈ *Pr*(X>Y) ≈ 0 (extreme stability) plausible?

- Pr(Y>X) ≈ Pr(X>Y) ≈ 0 means that changes are extremely unlikely within short time intervals such as those of known families
- Is this so? Given a set of variables, how many of them show changes within known families?

# Is $Pr$(Y>X) ≈ $Pr$(X>Y) ≈ 0 (extreme stability) plausible?

- The minimum number of attested changes $C$ for a variable $V$ with $k$ attested types (“levels”, “choices”) in a family $F$ is

  $\min(C_F) = k_F - 1$

  A family: A A A A A B B B A A, so $k_F = 2$
  Minimum change scenarios:
  *A > B in one branch, the rest stays, *or*
  *B > A in one branch, the rest stays

  Another family: A A C A A B B B A A, so $k_F = 3$
  Minimum change scenarios:
  *A → B in $F_1$, *A → C in $F_2$, A stays in $F_3$ or
  *B → A in $F_1$, *B → C in $F_2$, B stays in $F_3$, or
  *C → A in $F_1$, *C → B in $F_2$, C stays in $F_3$

  That’s the logical minima. (There can always be many more!)

# Is $Pr(Y>X) \approx Pr(X>Y) \approx 0$ (extreme stability) plausible?

## Is *Pr*(Y>X) ≈ *Pr*(X>Y) ≈ 0 (extreme stability) plausible?

- Test for each variable whether the **observed minimum of changes** per family **exceeds what can be expected** under some assumed probability of change $\pi$, *and no other factors.*

# Is $Pr$(Y>X) ≈ $Pr$(X>Y) ≈ 0 (extreme stability) plausible?

- Test for each variable whether the **observed minimum of changes** per family **exceeds what can be expected** under some assumed probability of change $\pi$, *and no other factors.*

- **Criterion of excess:** the proportion of min($C_F$) out of the total minimum of opportunities $O_F$ for change is unexpected for an assumed probability of change $\pi$ if the proportion exceeds the proportion under $H_0$ in a binomial test (at a 5% rejection level)

# Is $Pr$(Y>X) ≈ $Pr$(X>Y) ≈ 0 (extreme stability) plausible?

- Test for each variable whether the **observed minimum of changes** per family **exceeds what can be expected** under some assumed probability of change $\pi$, *and no other factors.*

- **Criterion of excess:** the proportion of min($C_F$) out of the total minimum of opportunities $O_F$ for change is unexpected for an assumed probability of change $\pi$ if the proportion exceeds the proportion under $H_0$ in a binomial test (at a 5% rejection level)

- Minimum opportunities for change min($O_F$) = $(k_V\text{-}1) \cdot N(\text{families})$ where $k_V$ is the number of types defined by a variable (what's possible), e.g. $k$=2, $N$=50 families: 50 opportunities for $V$ to change at least once
  $k$=3, $N$=50 families: 100 opportunities for $V$ to change at least once

# Is $Pr$(Y>X) ≈ $Pr$(X>Y) ≈ 0 (extreme stability) plausible?

# Is $Pr$(Y>X) ≈ $Pr$(X>Y) ≈ 0 (extreme stability) plausible?

- An example: assume probability of change is $\pi = .15$

# Is $Pr$(Y>X) ≈ $Pr$(X>Y) ≈ 0 (extreme stability) plausible?

- An example: assume probability of change is $\pi = .15$
  - if we find min($C_F$) = 20 out of min($O_F$) = 50, this is unexpected under $\pi = .15$ (at a 5% rejection level) → "unexpected"

# Is $Pr$(Y>X) ≈ $Pr$(X>Y) ≈ 0 (extreme stability) plausible?

- An example: assume probability of change is $\pi = .15$
  - if we find min($C_F$) = 20 out of min($O_F$) = 50, this is unexpected under $\pi = .15$ (at a 5% rejection level) → "unexpected"
  - if we find min($C_F$) = 20 out of min($O_F$) = 100, this is expected under $\pi = .15$ → "expected"

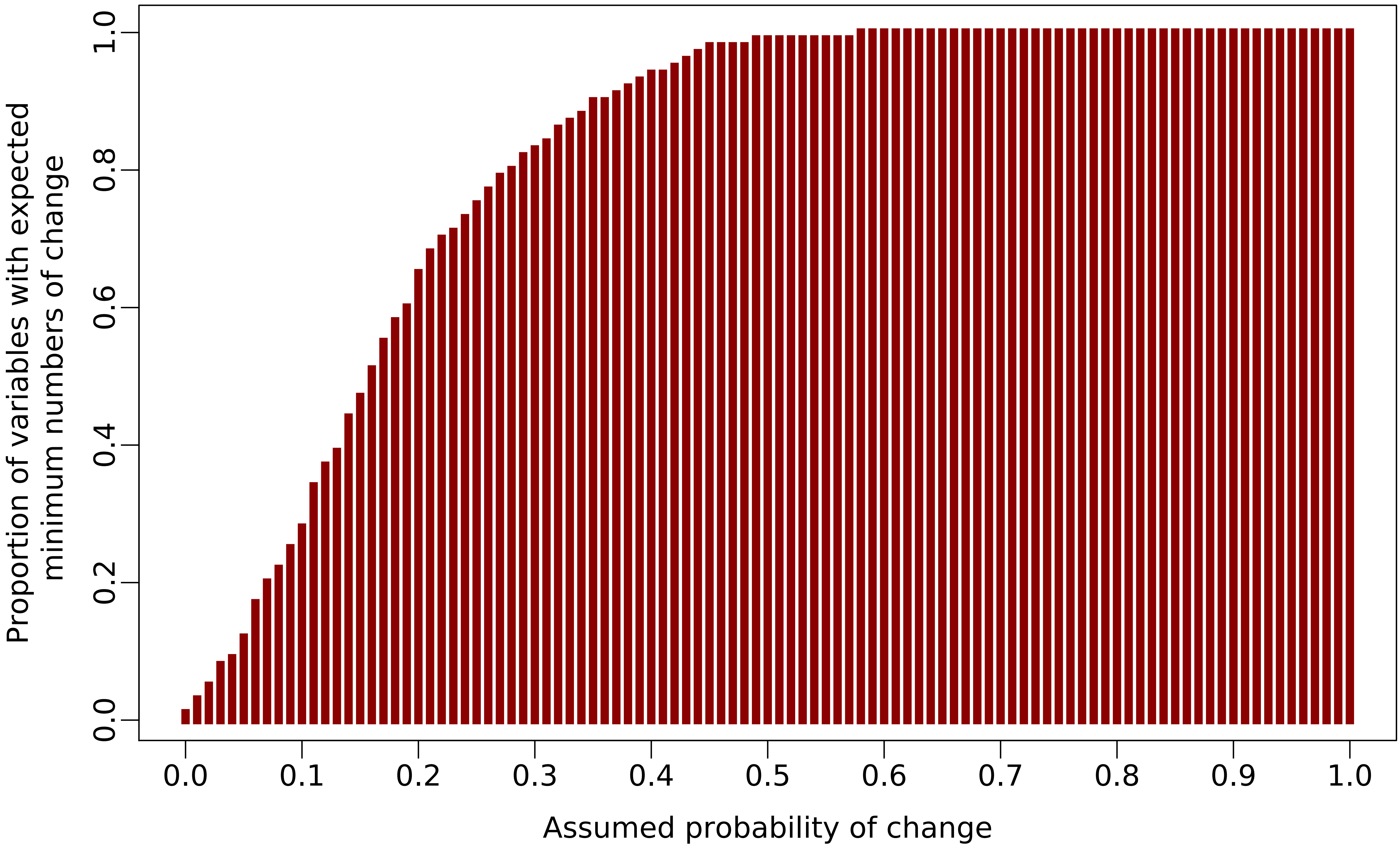
# Is $Pr$(Y>X) ≈ $Pr$(X>Y) ≈ 0 (extreme stability) plausible?

- An example: assume probability of change is $\pi = .15$
  - if we find min($C_F$) = 20 out of min($O_F$) = 50, this is unexpected under $\pi = .15$ (at a 5% rejection level) → "unexpected"
  - if we find min($C_F$) = 20 out of min($O_F$) = 100, this is expected under $\pi = .15$ → "expected"
- NB: since we only look at minima, this underestimates the number of unexpected changes, i.e. it favors small $\pi$!

# Is $Pr$(Y>X) ≈ $Pr$(X>Y) ≈ 0 (extreme stability) plausible?

- An example: assume probability of change is $\pi = .15$
  - if we find min($C_F$) = 20 out of min($O_F$) = 50, this is unexpected under $\pi = .15$ (at a 5% rejection level) → "unexpected"
  - if we find min($C_F$) = 20 out of min($O_F$) = 100, this is expected under $\pi = .15$ → "expected"
- NB: since we only look at minima, this underestimates the number of unexpected changes, i.e. it favors small $\pi$!
- Compute the proportion of variables for which min($C_F$) is expected, given the assumption of a specific value of $\pi$ between 0 and 1

# Is $Pr(Y>X) \approx Pr(X>Y) \approx 0$ (extreme stability) plausible?




448 variables from WALS and AUTOTYP

# Is $Pr$(Y>X) ≈ $Pr$(X>Y) ≈ 0 (extreme stability) plausible?

- Some of the variables with min($C_F$) expected under $\pi = .01$:

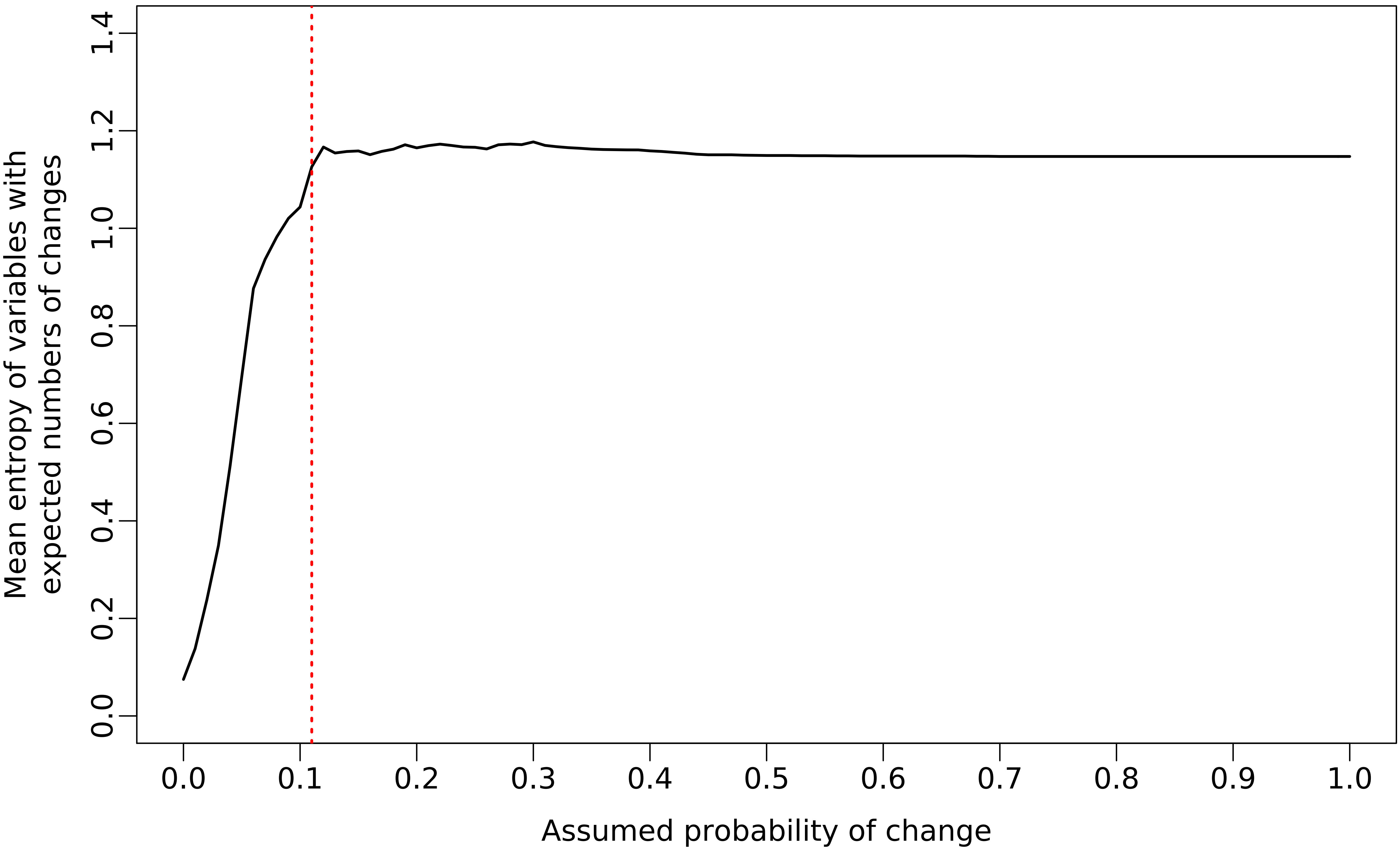
# Is $Pr(Y>X) \approx Pr(X>Y) \approx 0$ (extreme stability) plausible?

- Some of the variables with $\min(C_F)$ expected under $\pi = .01$:

| Variable (and data source) | Changes $N_{min}$ | Opportunities $N_{min}$ | Entropy $\hat{H}$ | Ratio of values |
|---|---|---|---|---|
| Interrog./decl. distinction (Dryer, 2005a) | 1 | 89 | 0.01 | 841:1 |
| Indep. subject pronouns (Daniel, 2005) | 0 | 31 | 0.07 | 258:2 |
| Tonal case (autotyp and Dryer, 2005b) | 3 | 91 | 0.07 | 698:6 |
| Stem flexivity condit. by NEG (autotyp) | 0 | 40 | 0.12 | 141:1:1 |
| ‘Have’-perfect (Dahl & Velupillai, 2005) | 1 | 15 | 0.35 | 101:7 |
| Co-exponent type of NEG (autotyp) | 4 | 234 | 0.60 | 185:5:3:1:1:1:1:1:1:1 |

- This is typical: $\pi \leq .10$ suggest **rara vs. universalia** distributions, not extreme stability

# Is $Pr(Y>X) \approx Pr(X>Y) \approx 0$ (extreme stability) plausible?




Bickel in press ("Distributional biases in language families"), in Fs. Nichols

# Interim Summary

# Interim Summary

- So, the key assumption behind the Family Bias Method — synchronic biases reflect directional trends in diachrony — is justified by exclusion of the alternatives: (i) deny the synchronic bias; (ii) assume extreme stability

# Interim Summary

- So, the key assumption behind the Family Bias Method — synchronic biases reflect directional trends in diachrony — is justified by exclusion of the alternatives: (i) deny the synchronic bias; (ii) assume extreme stability
- But how to implement the Family Bias Method?

# Three things we need

# Three things we need

1. Groups of demonstrably related languages: families, established by the Comparative Method.

   This guarantees that the current distribution descends from a single common ancestor via processes of innovation and retention

# Three things we need

1. Groups of demonstrably related languages: families, established by the Comparative Method.

   This guarantees that the current distribution descends from a single common ancestor via processes of innovation and retention

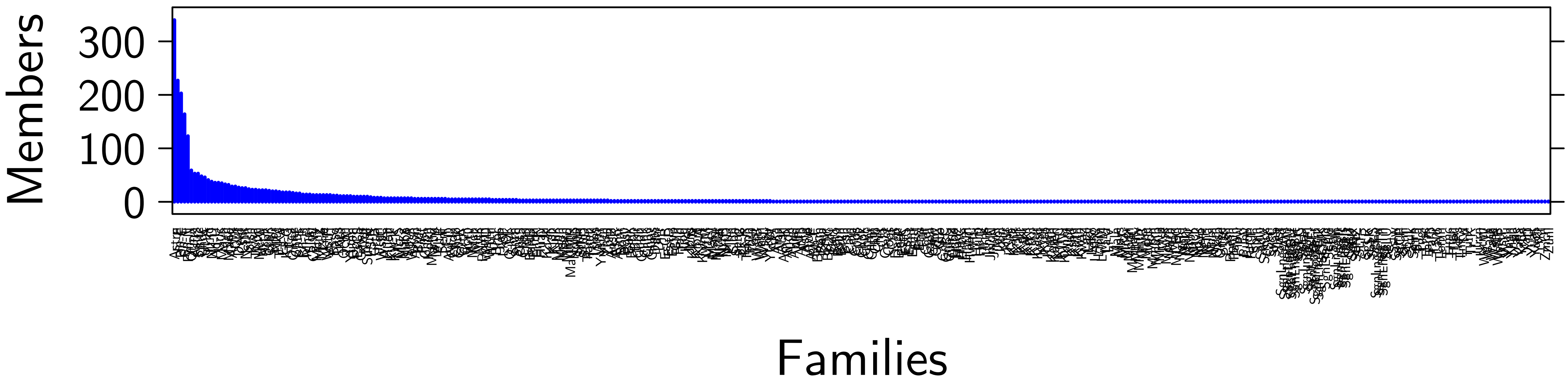
2. A way of evaluating synchronic preferences as indicators of diachronic biases

1. Groups of demonstrably related languages: families, established by the Comparative Method.

   This guarantees that the current distribution descends from a single common ancestor via processes of innovation and retention

2. A way of evaluating synchronic preferences as indicators of diachronic biases

3. A way of dealing with small families and isolates:



- Several options, e.g.

- Several options, e.g.
  - binomial tests for binomial variables, and then generalize to multinomial designs (currently the only option we have fully implemented)

    R package `familybias` available

    at `http://www.uzh.ch/spw/software`

## Detecting diachronic biases

- Several options, e.g.
  - binomial tests for binomial variables, and then generalize to multinomial designs (currently the only option we have fully implemented)

    R package `familybias` available

    at `http://www.uzh.ch/spw/software`
  - estimate likelihoods of synchronic distributions given diachronic biases (work in progress)

- Several options, e.g.
  - binomial tests for binomial variables, and then generalize to multinomial designs (currently the only option we have fully implemented)

    R package `familybias` available

    at `http://www.uzh.ch/spw/software`
  - estimate likelihoods of synchronic distributions given diachronic biases (work in progress)
- Justification of the binomial test approach by computer simulation (joint work with Taras Zakharko)

## Justification of binomial tests for detecting diachronic biases

Simulation of a discrete-time Markov process, where language varieties can (within steps of ca. 100 years $\sim$ 3 generations)

- give birth: Poisson process with **birth rate $\boldsymbol{\lambda}$** within [.7, .9]

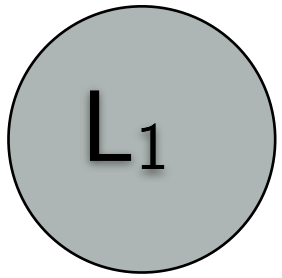
  meaning that it takes 1 or 2 steps (100-200 years, 3-6 generation) for a new language variety to get established, *on average*

- die or stay live: Bernoulli process with **survival prob. $\boldsymbol{\pi}$** within [.1, .2]

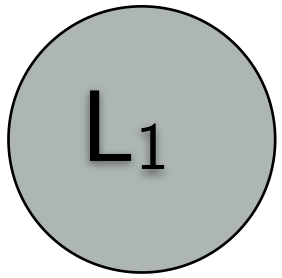
  meaning that most varieties die after 1 or 2 steps (100-200 years), *on average*

(for simplicity, $\lambda$ and $\pi$ are assumed to be constant within one simulation)

A proto-language, $t = 0$



**Step 1 (about 100y or 3 generations)**

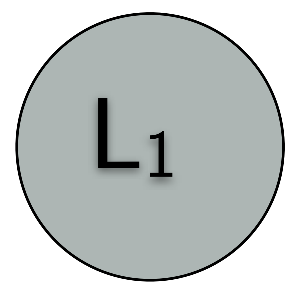


# Simulating birth and survival: an example

**Step 1 (about 100y or 3 generations)**

$\texttt{rpois}(.8) \;= 1$

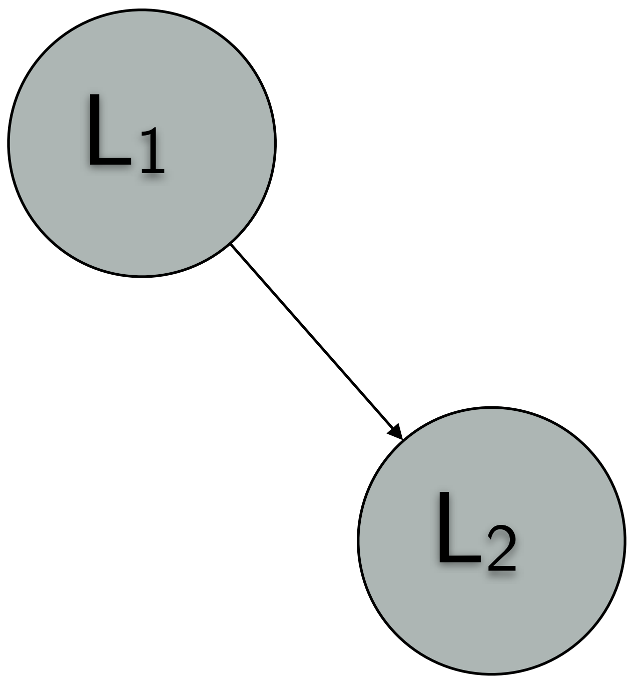
$\texttt{rbinom}(.1) = 1$



**Step 1 (about 100y or 3 generations)**

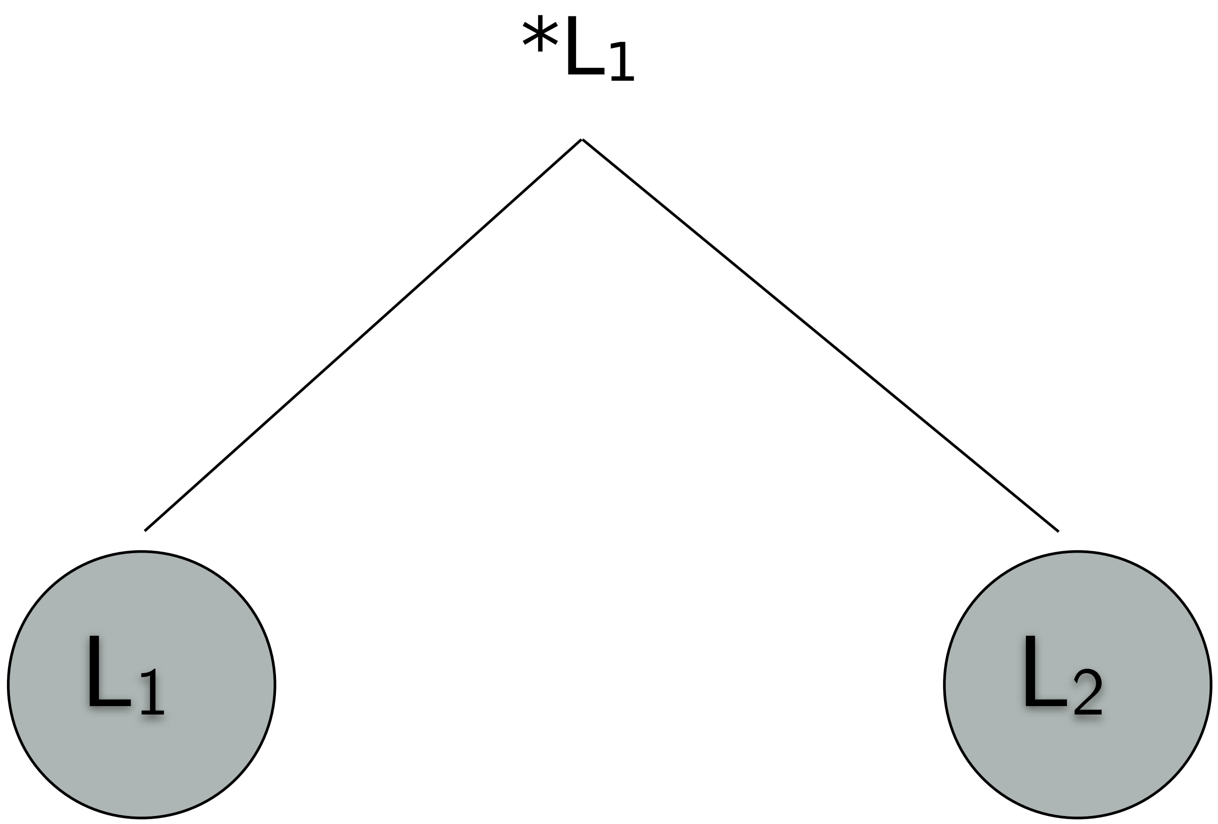
$\texttt{rpois}(.8) = 1$

$\texttt{rbinom}(.1) = 1$



Result after one step, $t = 100$y

## Result after one step, $t = 100$y




(conservative variety of $L_1$, no or negligible changes)

(innovative variety of $L_1$, coexisting with it)

**Step 2 (another 100y or 3 generations)**

**Step 2 (another 100y or 3 generations)**

$\texttt{rpois}(.8) = 0$ $\quad$ $\texttt{rpois}(.8) = 2$

$\texttt{rbinom}(.1) = 0$ $\quad$ $\texttt{rbinom}(.1) = 0$

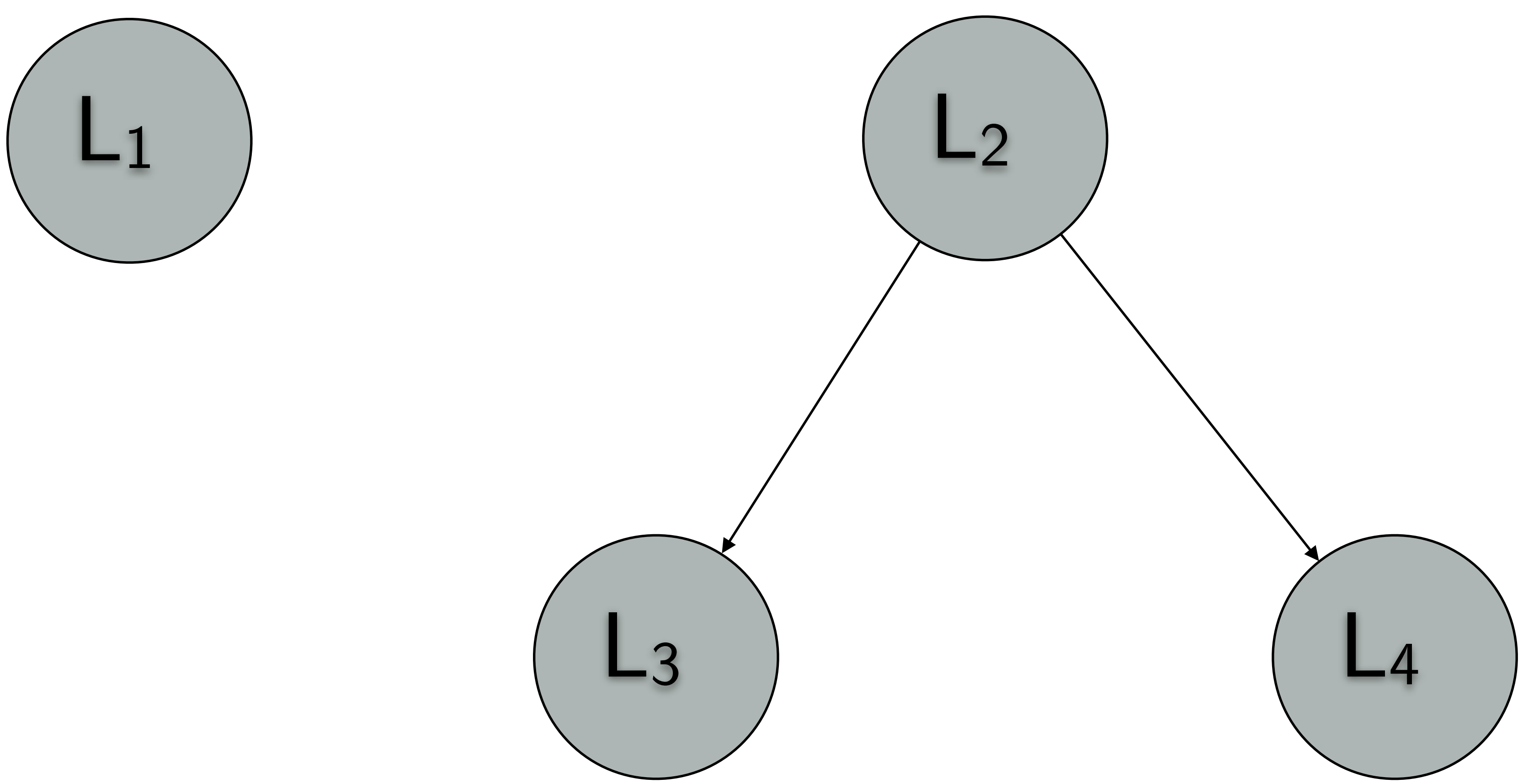
**Step 2 (another 100y or 3 generations)**

$\texttt{rpois}(.8) = 0$

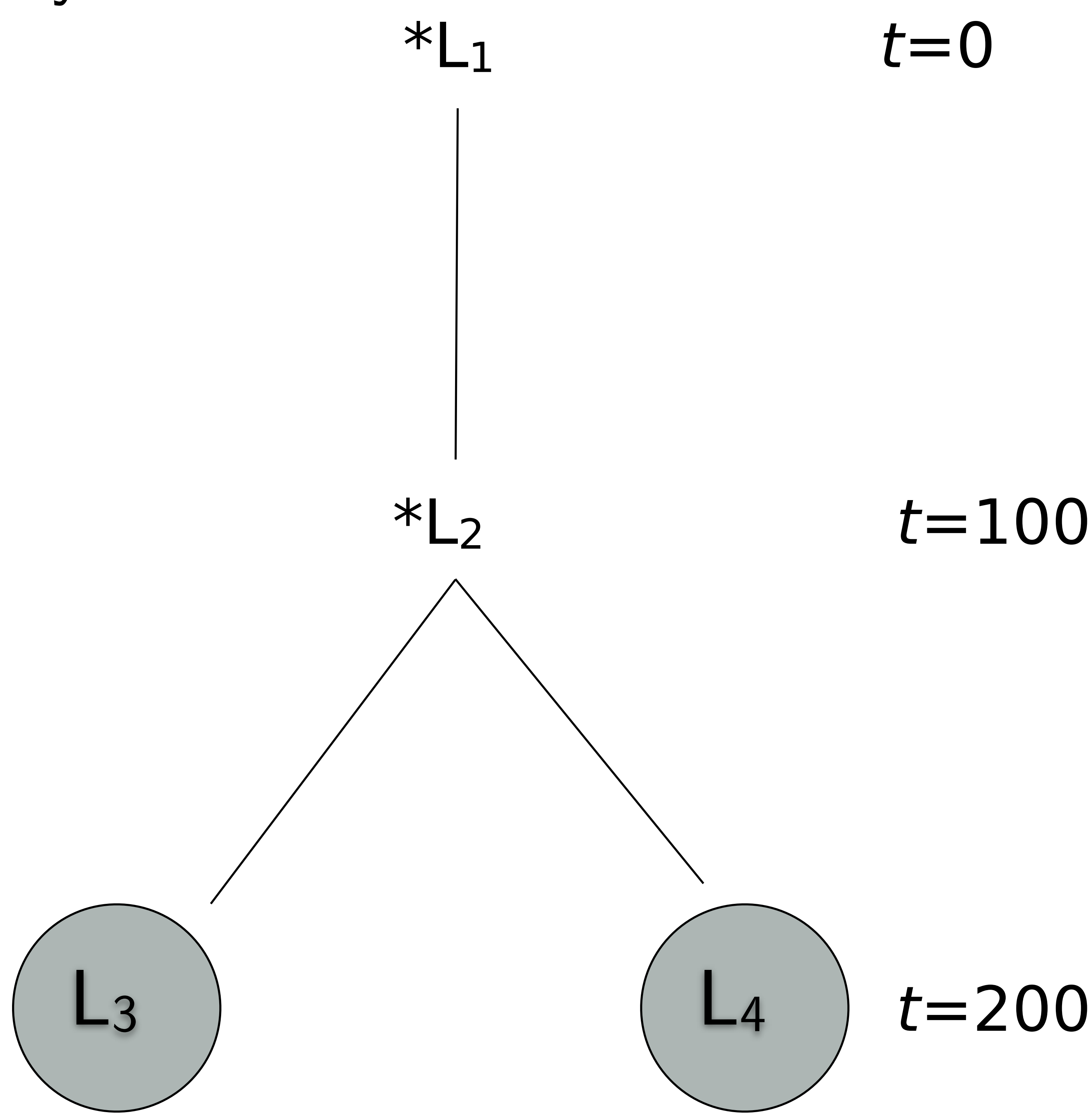
$\texttt{rbinom}(.1) = 0$

$\texttt{rpois}(.8) = 2$

$\texttt{rbinom}(.1) = 0$

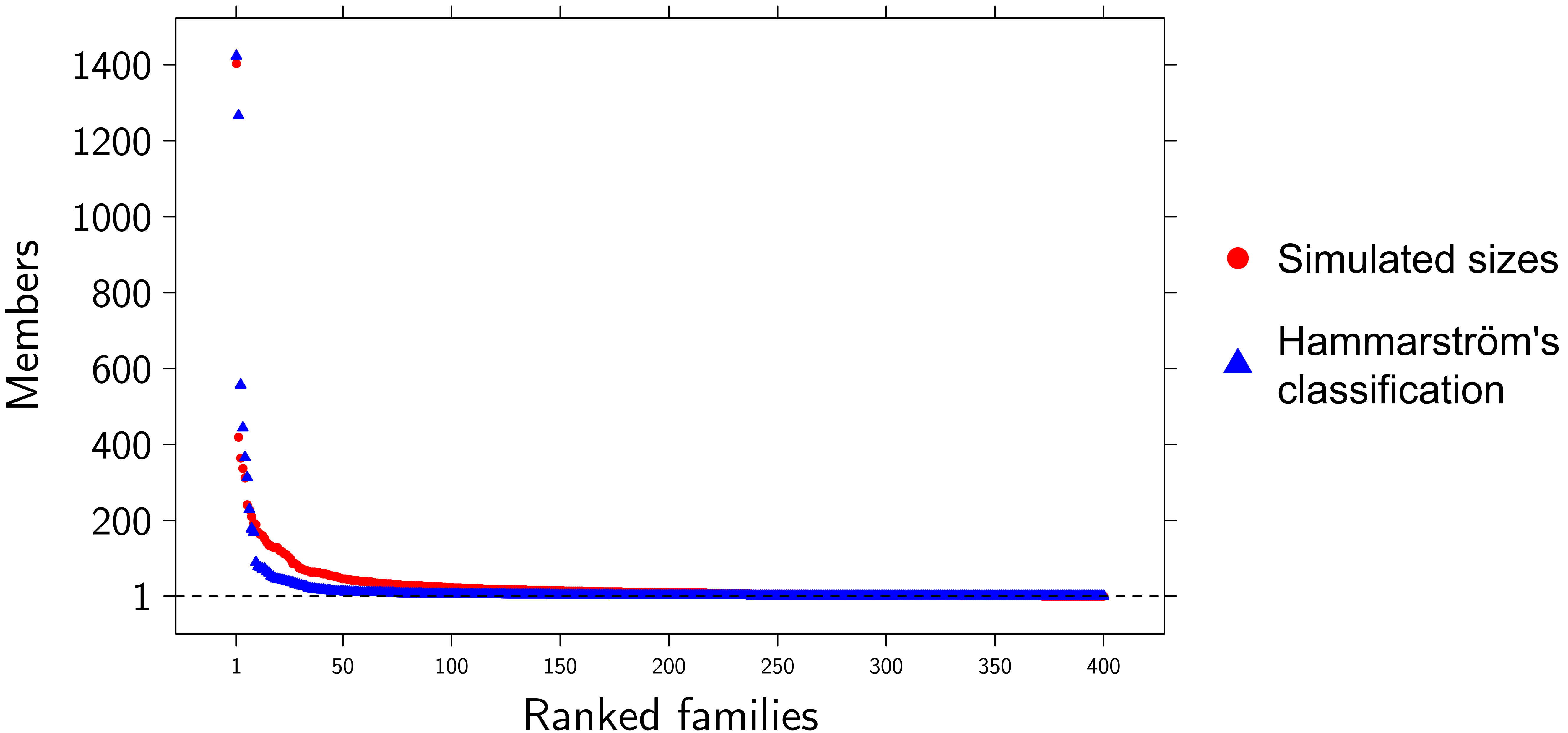


Result after two steps, $t = 200$y



# Simulating birth and survival: reality check

400 simulated families with randomly chosen birth rates $\lambda$ between [.7,.8] and survival probabilities $\pi$ between [.1,.2], running randomly between 30 and 50 steps, i.e. 3'000 - 5'000y:



# Simulating change in this model

# Simulating change in this model

- A language is defined (as usual) by idiosyncracies (Saussurian signs), and these are associated with typological variables

$$\mathrm{L} = \{\mathrm{I}_1, \mathrm{I}_2, \mathrm{I}_3 \ldots\} \rightarrow \{\mathrm{V}_1, \mathrm{V}_2, \mathrm{V}_3 \ldots\}$$

# Simulating change in this model

- A language is defined (as usual) by idiosyncracies (Saussurian signs), and these are associated with typological variables

$$L = \{I_1, I_2, I_3 \ldots\} \rightarrow \{V_1, V_2, V_3 \ldots\}$$

- any *V* can change whenever some *I* change, i.e. when a language is born

# Simulating change in this model

- A language is defined (as usual) by idiosyncracies (Saussurian signs), and these are associated with typological variables

  $L = \{I_1, I_2, I_3 \ldots\} \rightarrow \{V_1, V_2, V_3 \ldots\}$

- any *V* can change whenever some *I* change, i.e. when a language is born
- simulate behavior of *V* during birth:

## Simulating change in this model

- A language is defined (as usual) by idiosyncracies (Saussurian signs), and these are associated with typological variables

$$L = \{I_1, I_2, I_3 \dots\} \rightarrow \{V_1, V_2, V_3 \dots\}$$

- any *V* can change whenever some *I* change, i.e. when a language is born
- simulate behavior of *V* during birth:
  - without a diachronic bias:

    $Pr(Y>X) \approx Pr(X>Y)$ for some binomial variable *V*

# Simulating change in this model

- A language is defined (as usual) by idiosyncracies (Saussurian signs), and these are associated with typological variables

  $L = \{I_1, I_2, I_3 \dots \} \rightarrow \{V_1, V_2, V_3 \dots\}$

- any *V* can change whenever some *I* change, i.e. when a language is born
- simulate behavior of *V* during birth:
  - without a diachronic bias:

    $Pr(Y>X) \approx Pr(X>Y)$ for some binomial variable *V*

  - with a diachronic bias:

    $Pr(Y>X) > Pr(X>Y)$ for some binomial variable *V*

# Simulating change in this model

- A language is defined (as usual) by idiosyncracies (Saussurian signs), and these are associated with typological variables

  $L = \{I_1, I_2, I_3 \ldots\} \rightarrow \{V_1, V_2, V_3 \ldots\}$

- any *V* can change whenever some *I* change, i.e. when a language is born
- simulate behavior of *V* during birth:
  - without a diachronic bias:

    $Pr(Y>X) \approx Pr(X>Y)$ for some binomial variable *V*

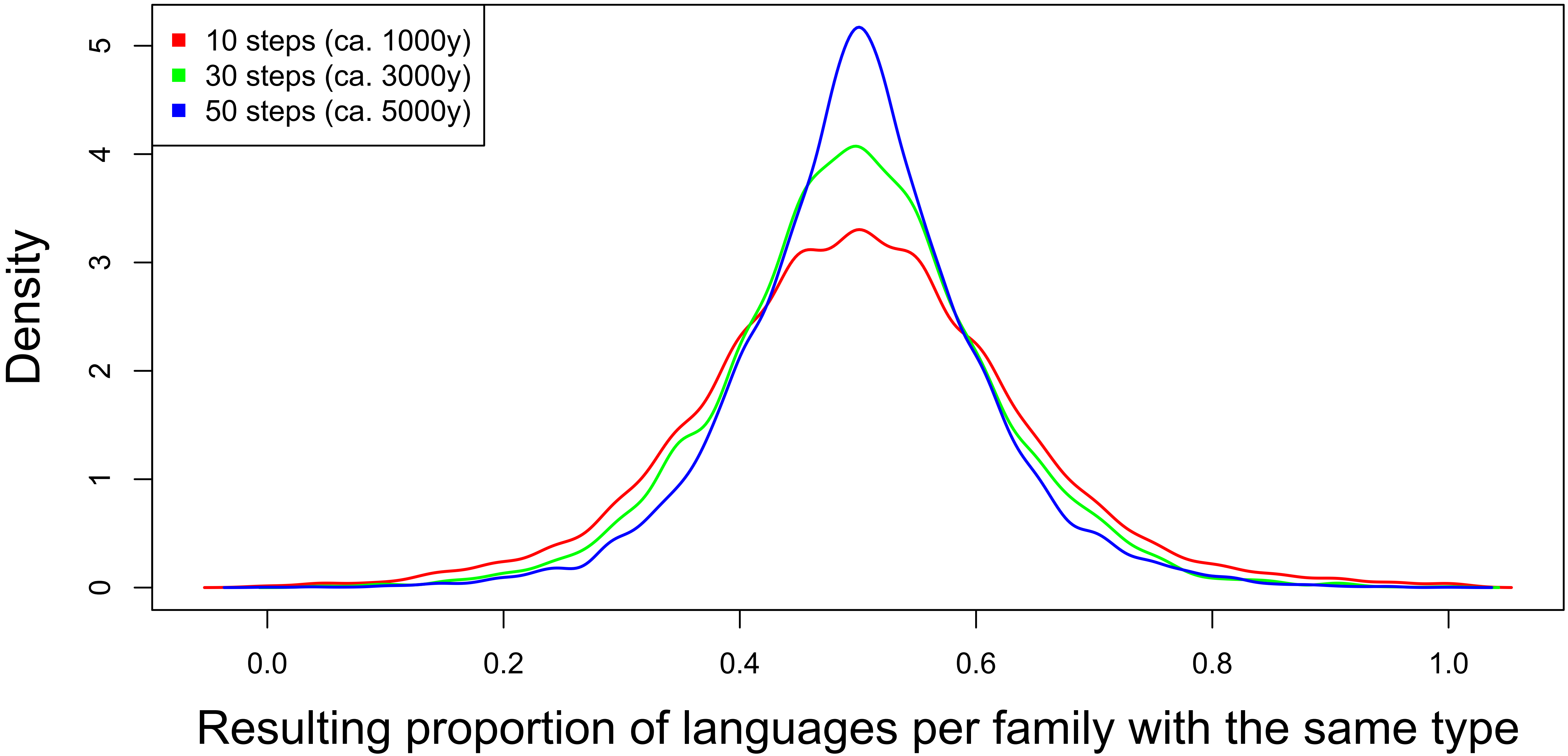
  - with a diachronic bias:

    $Pr(Y>X) > Pr(X>Y)$ for some binomial variable *V*

- and examine the resulting distribution in families that have at least 20 survivors in the simulations (10k runs)

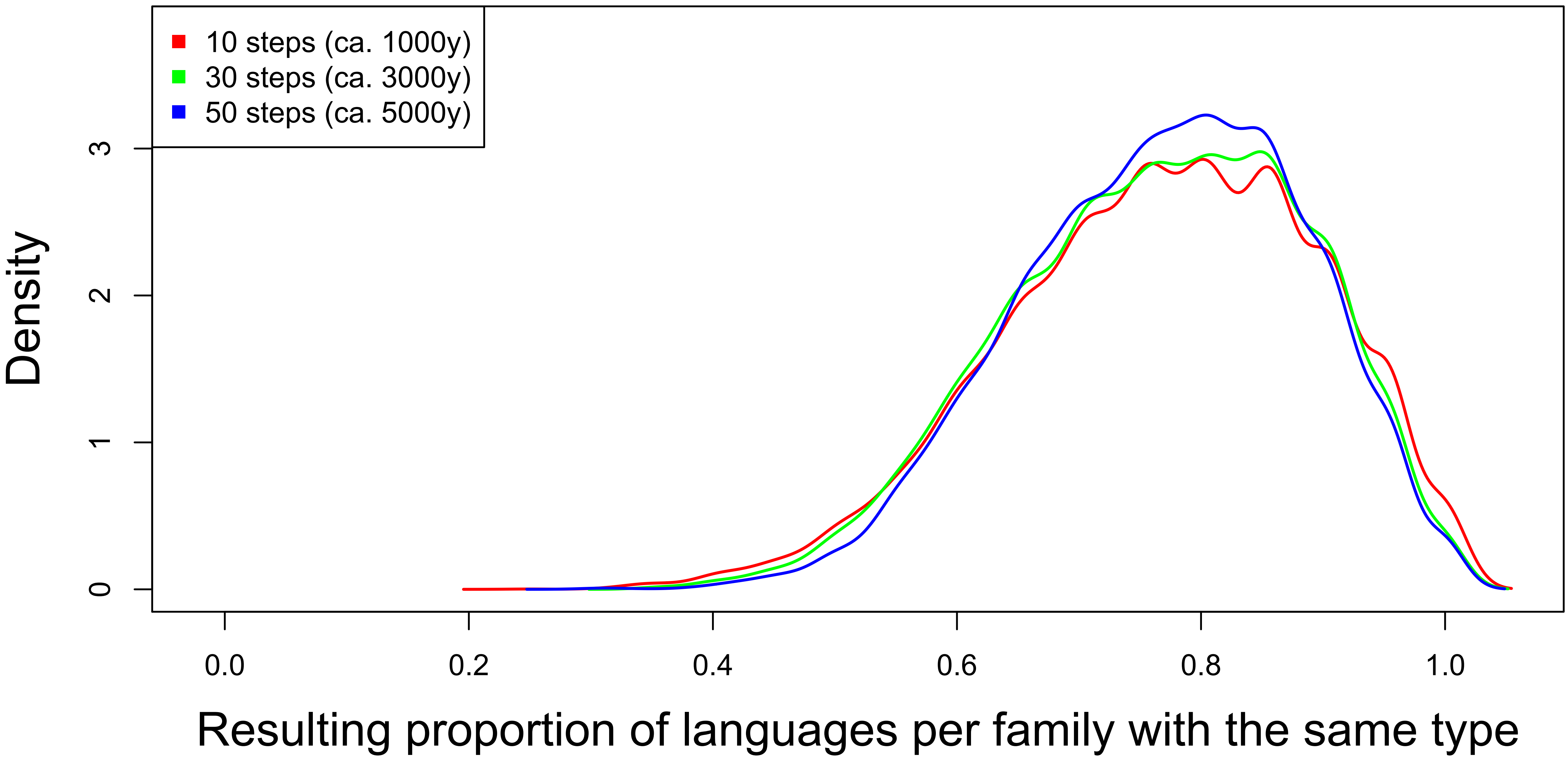
# Simulating change in this model

- without bias, |Pr(Y>X) - Pr(X>Y)| ≤ .05



# Simulating change in this model

- with a bias, |Pr(Y>X) - Pr(X>Y)| ≥ .25



- The clear shift in the probability mass suggests that an exact binomial test (with a 10% rejection level) is a reasonable bias test (families with at least 20 members, 10k simulations):

| | no bias detected | bias detected |
|---|---|---|
| family has no bias | 0.87 | 0.13 |
| family has bias | 0.19 | 0.81 |

False positives

False negatives

# Extrapolations to small families and isolates

- Assumption: *An isolate is the sole survivor of a larger unknown (perhaps dead) family and its development is not in principle different from that of known families*

# Extrapolations to small families and isolates

- Assumption: *An isolate is the sole survivor of a larger unknown (perhaps dead) family and its development is not in principle different from that of known families*
- Use all information we have about known families to estimate what kinds of trends there are in unknown families:

# Extrapolations to small families and isolates

- Assumption: *An isolate is the sole survivor of a larger unknown (perhaps dead) family and its development is not in principle different from that of known families*

- Use all information we have about known families to estimate what kinds of trends there are in unknown families:

  - the range of attested values (e.g. X bias, Y bias, diverse), with $H_0$ probabilities of ⅓ each

# Extrapolations to small families and isolates

- Assumption: *An isolate is the sole survivor of a larger unknown (perhaps dead) family and its development is not in principle different from that of known families*
- Use all information we have about known families to estimate what kinds of trends there are in unknown families:
  - the range of attested values (e.g. X bias, Y bias, diverse), with $H_0$ probabilities of ⅓ each
  - the actual values in small families

# Extrapolations to small families and isolates

- Assumption: *An isolate is the sole survivor of a larger unknown (perhaps dead) family and its development is not in principle different from that of known families*

- Use all information we have about known families to estimate what kinds of trends there are in unknown families:

  - the range of attested values (e.g. X bias, Y bias, diverse), with $H_0$ probabilities of ⅓ each

  - the actual values in small families

  - the proportion of biased vs. diverse large families

# Extrapolations to small families and isolates

- Assumption: *An isolate is the sole survivor of a larger unknown (perhaps dead) family and its development is not in principle different from that of known families*
- Use all information we have about known families to estimate what kinds of trends there are in unknown families:
  - the range of attested values (e.g. X bias, Y bias, diverse), with $H_0$ probabilities of ⅓ each
  - the actual values in small families
  - the proportion of biased vs. diverse large families
- Various techniques for extrapolation. One technique:

# Extrapolations to small families and isolates

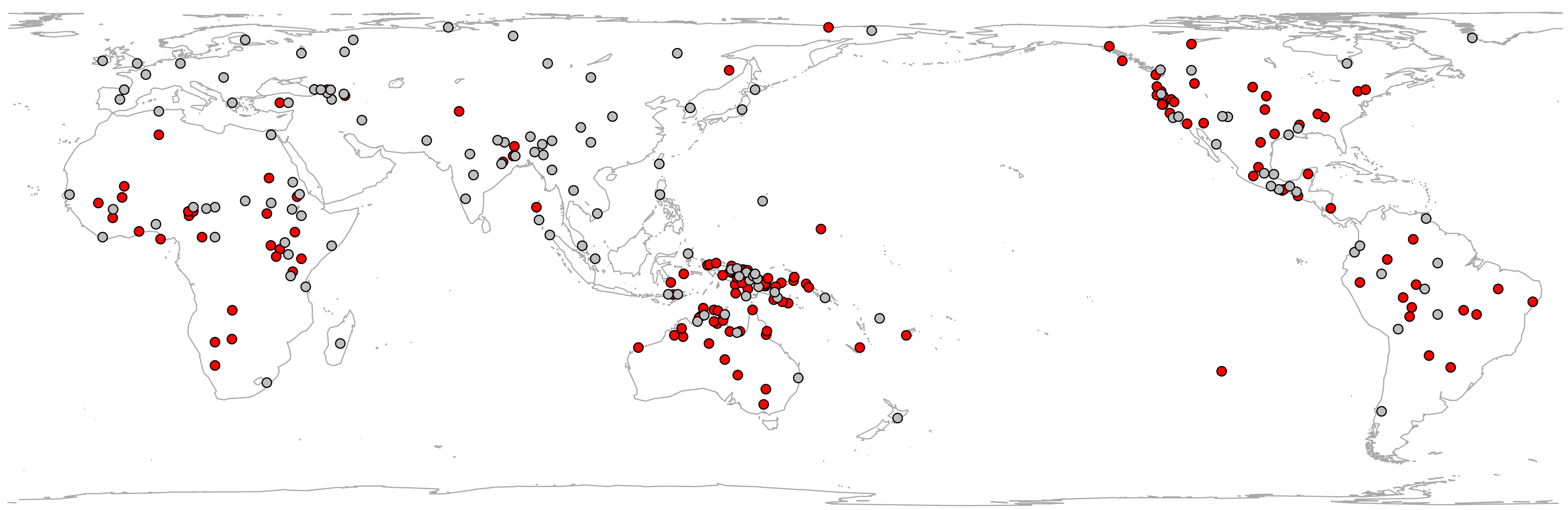
1. Estimate the proportion *Pr(bias)* of small families on the basis of what we know from large families (using Laplace's Rule of Succession: if 7/8 large families are biased, assume 8/10 smal families to be biased as well, no matter in what direction)

# Extrapolations to small families and isolates

1. Estimate the proportion *Pr(bias)* of small families on the basis of what we know from large families (using Laplace's Rule of Succession: if 7/8 large families are biased, assume 8/10 smal families to be biased as well, no matter in what direction)

   E.g. families with biases towards possessive classes (176 families, 274 languages)

| | Pr(bias) |
|---|---|
| Eurasia | 0.67 |
| Other | 0.40 |

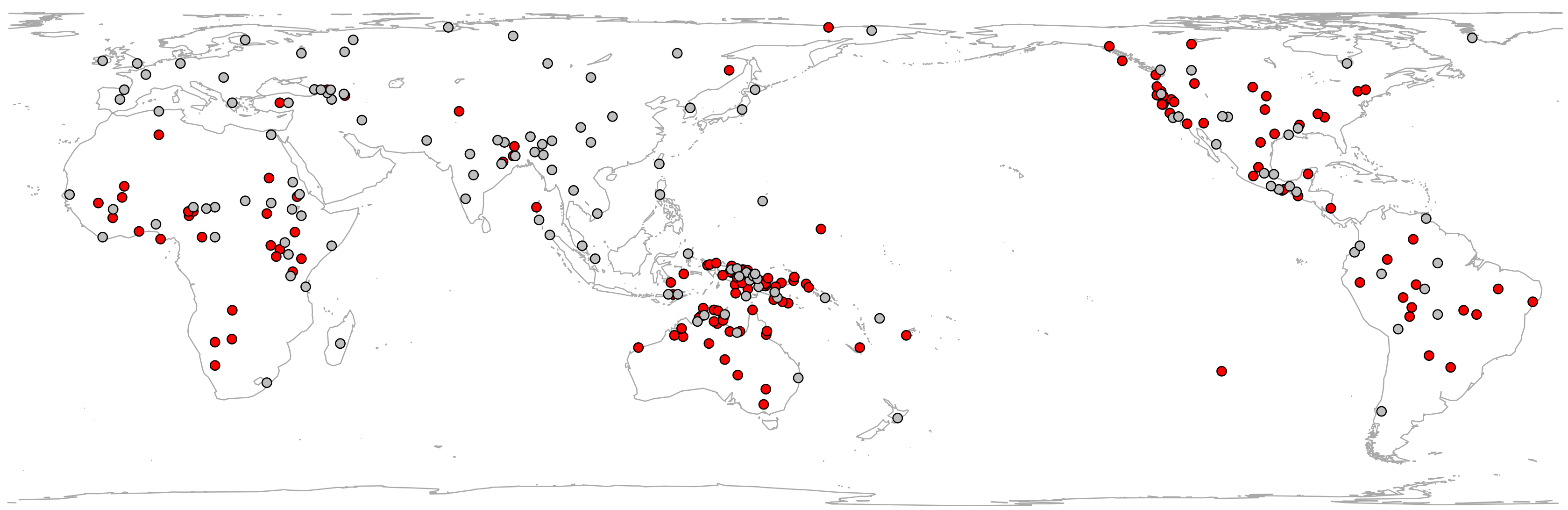


# Extrapolations to small families and isolates

1. Estimate the proportion *Pr(bias)* of small families on the basis of what we know from large families (using Laplace's Rule of Succession: if 7/8 large families are biased, assume 8/10 smal families to be biased as well, no matter in what direction)

   E.g. families with biases towards possessive classes (176 families, 274 languages)

| | Pr(bias) |
|---|---|
| Eurasia | 0.67 |
| Other | 0.40 |



→ Randomly take *Pr(bias)* small families and declare them has being the sole survivors of larger families with a bias, and 1-Pr(bias) as being the sole survivors of larger families without a bias

# Extrapolations to small families and isolates

# Extrapolations to small families and isolates

2. For those small families that are now assumed to be the sole survivor(s) of families with a bias, determine the direction of the bias:

2. For those small families that are now assumed to be the sole survivor(s) of families with a bias, determine the direction of the bias:

   1. estimate Pr(type is representative) from the strength of the bias in large families, e.g.

| | Pr(type is representative) |
|---|---|
| Eurasia | 0.92 |
| Other | 0.82 |

2. For those small families that are now assumed to be the sole survivor(s) of families with a bias, determine the direction of the bias:

   1. estimate Pr(type is representative) from the strength of the bias in large families, e.g.

| | Pr(type is representative) |
|---|---|
| Eurasia | 0.92 |
| Other | 0.82 |

   2. then

2. For those small families that are now assumed to be the sole survivor(s) of families with a bias, determine the direction of the bias:

    1. estimate Pr(type is representative) from the strength of the bias in large families, e.g.

| | Pr(type is representative) |
|---|---|
| Eurasia | 0.92 |
| Other | 0.82 |

    2. then

- with Pr(type is representative), take the type of the survivor(s) as the "real" direction of the bias

2. For those small families that are now assumed to be the sole survivor(s) of families with a bias, determine the direction of the bias:

   1. estimate Pr(type is representative) from the strength of the bias in large families, e.g.

| | Pr(type is representative) |
|---|---|
| Eurasia | 0.92 |
| Other | 0.82 |

   2. then

- with Pr(type is representative), take the type of the survivor(s) as the "real" direction of the bias
- with 1-Pr(type is representative), i.e. pick the only available or a random alternative type as the "real" direction of the bias

# Extrapolations to small families and isolates

## Extrapolations to small families and isolates

- Now these extrapolations use random assignments in three places:

## Extrapolations to small families and isolates

- Now these extrapolations use random assignments in three places:
  - when choosing which small families are 'biased' vs. 'diverse' (we know the proportion, but we don't know which ones they are)

## Extrapolations to small families and isolates

- Now these extrapolations use random assignments in three places:
  - when choosing which small families are 'biased' vs. 'diverse' (we know the proportion, but we don't know which ones they are)
  - when correcting for the possibility that the sole survivors may not be representative of their families, i.e. exceptions

## Extrapolations to small families and isolates

- Now these extrapolations use random assignments in three places:
  - when choosing which small families are ‘biased’ vs. ‘diverse’ (we know the proportion, but we don’t know which ones they are)
  - when correcting for the possibility that the sole survivors may not be representative of their families, i.e. exceptions
  - when picking a value for small families estimated to be survivors of biased families (if the small family is not uniform)

## Extrapolations to small families and isolates

- Now these extrapolations use random assignments in three places:
  - when choosing which small families are 'biased' vs. 'diverse' (we know the proportion, but we don't know which ones they are)
  - when correcting for the possibility that the sole survivors may not be representative of their families, i.e. exceptions
  - when picking a value for small families estimated to be survivors of biased families (if the small family is not uniform)

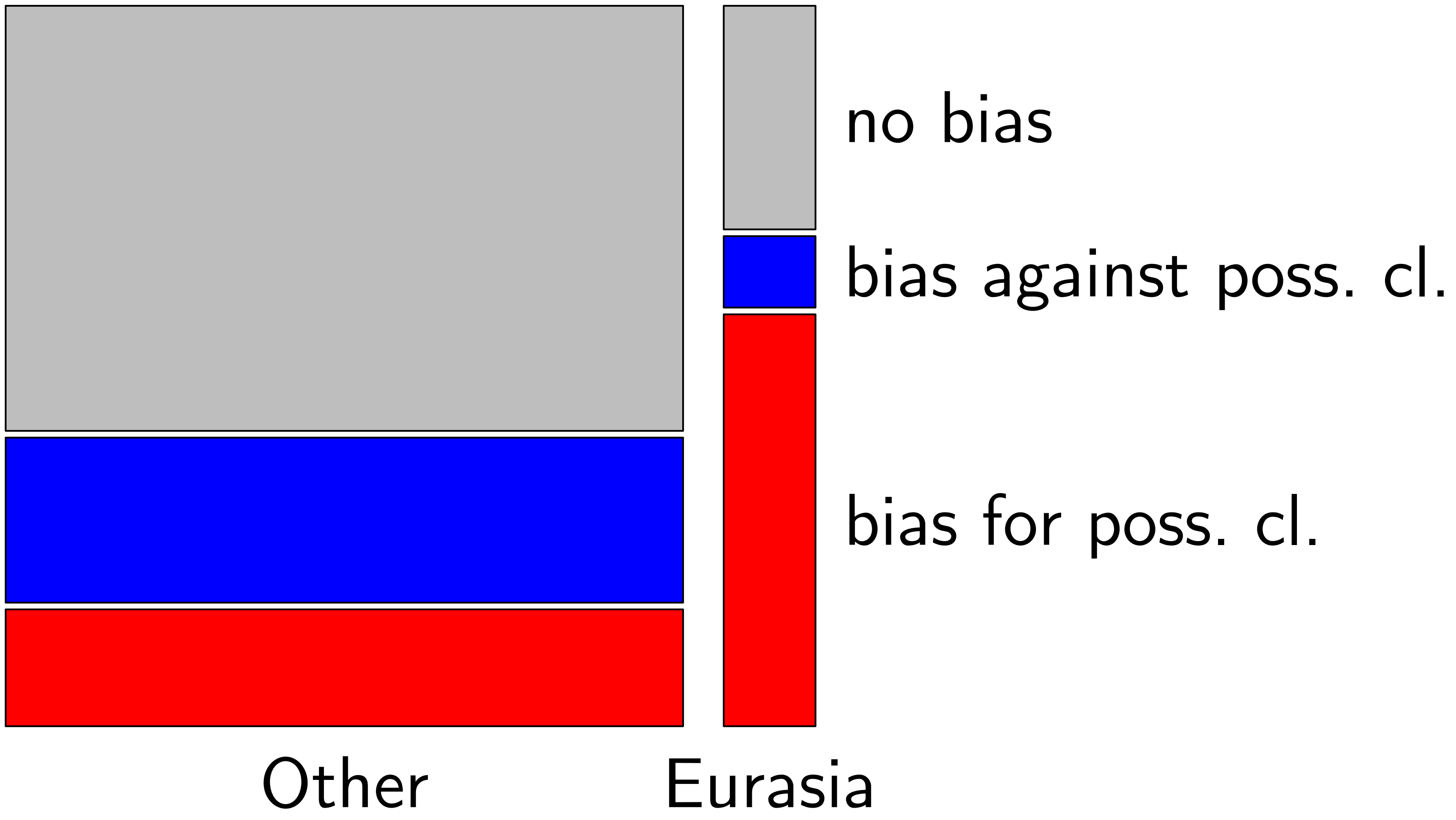
  These random assignments introduce a statistical error but this can be assumed to be normally distributed

- Now these extrapolations use random assignments in three places:
  - when choosing which small families are 'biased' vs. 'diverse' (we know the proportion, but we don't know which ones they are)
  - when correcting for the possibility that the sole survivors may not be representative of their families, i.e. exceptions
  - when picking a value for small families estimated to be survivors of biased families (if the small family is not uniform)

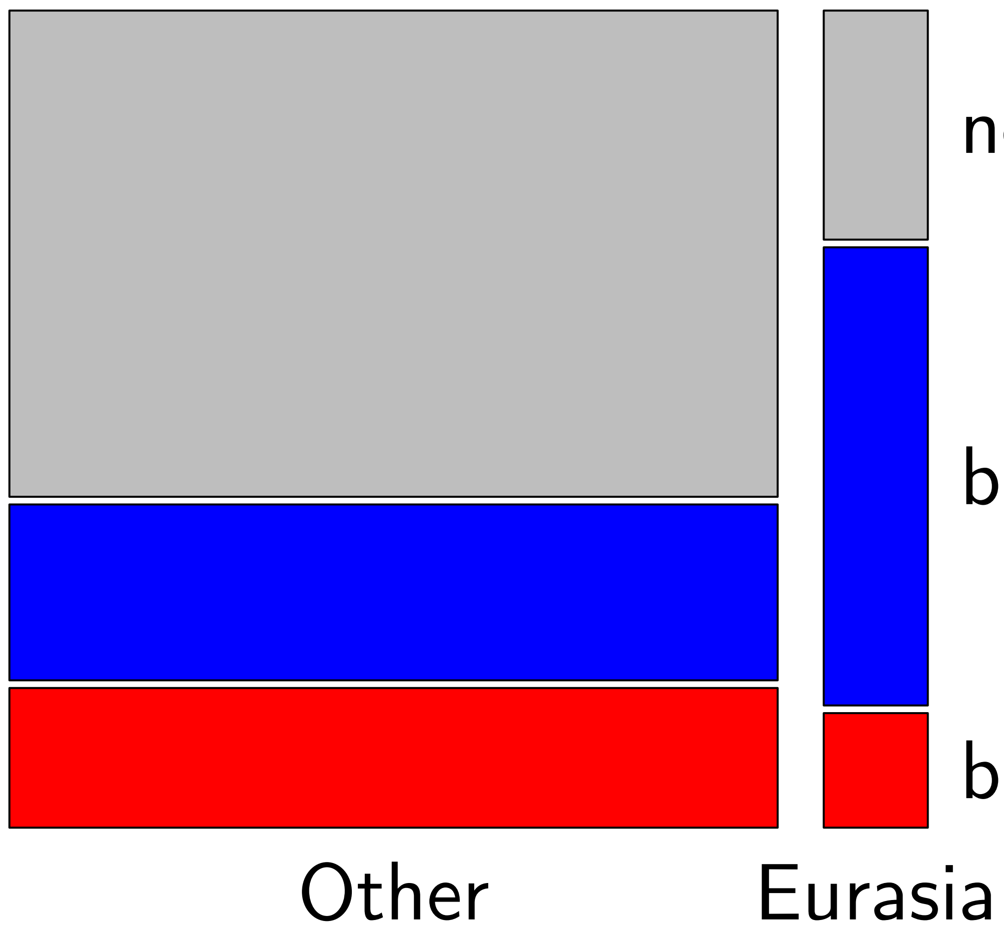
These random assignments introduce a statistical error but this can be assumed to be normally distributed

$\rightarrow$ Therefore, we can take the mean of a set of random assignments, e.g. the mean of 2,000 extrapolations

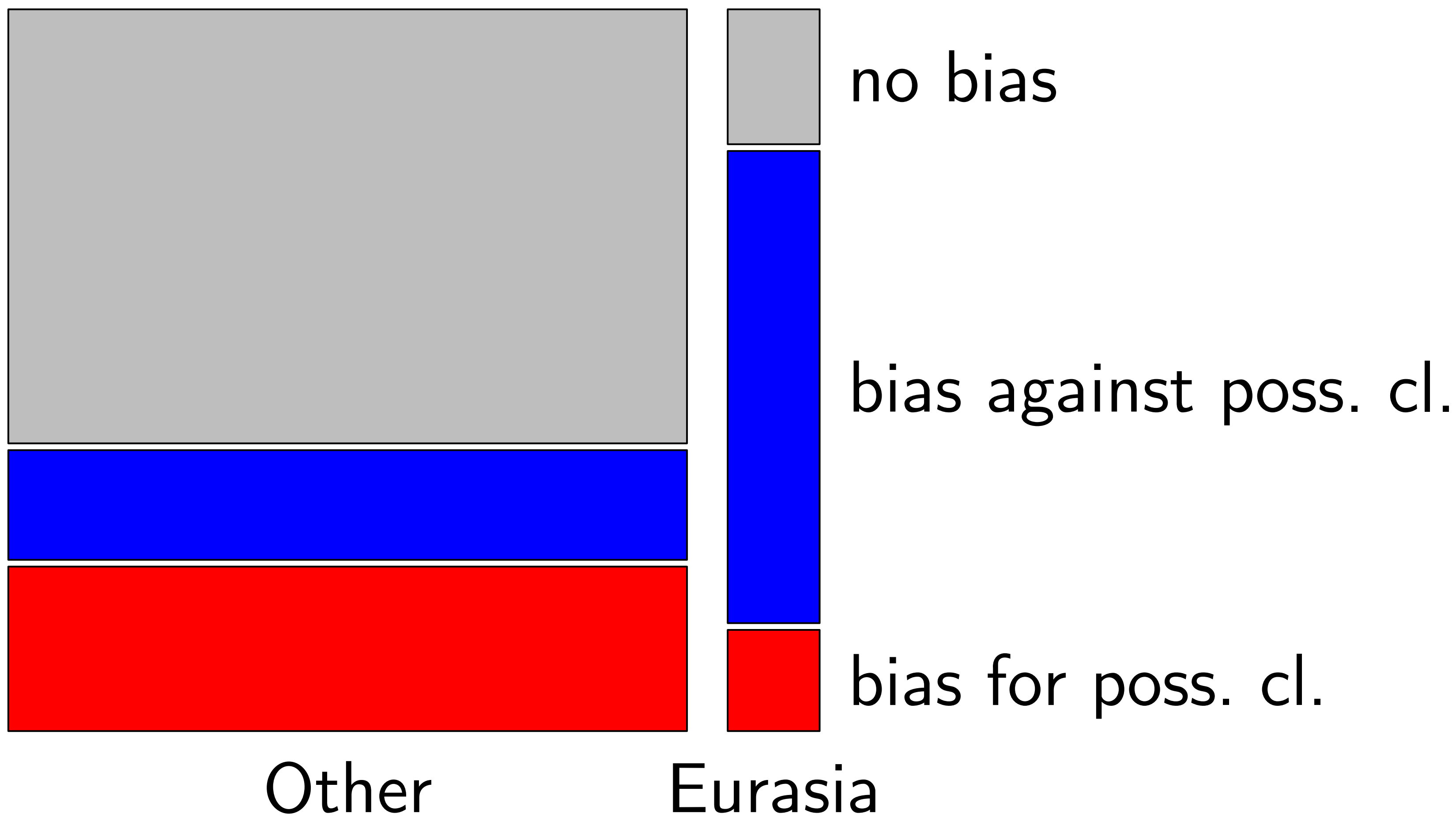
# Extrapolations to small families and isolates



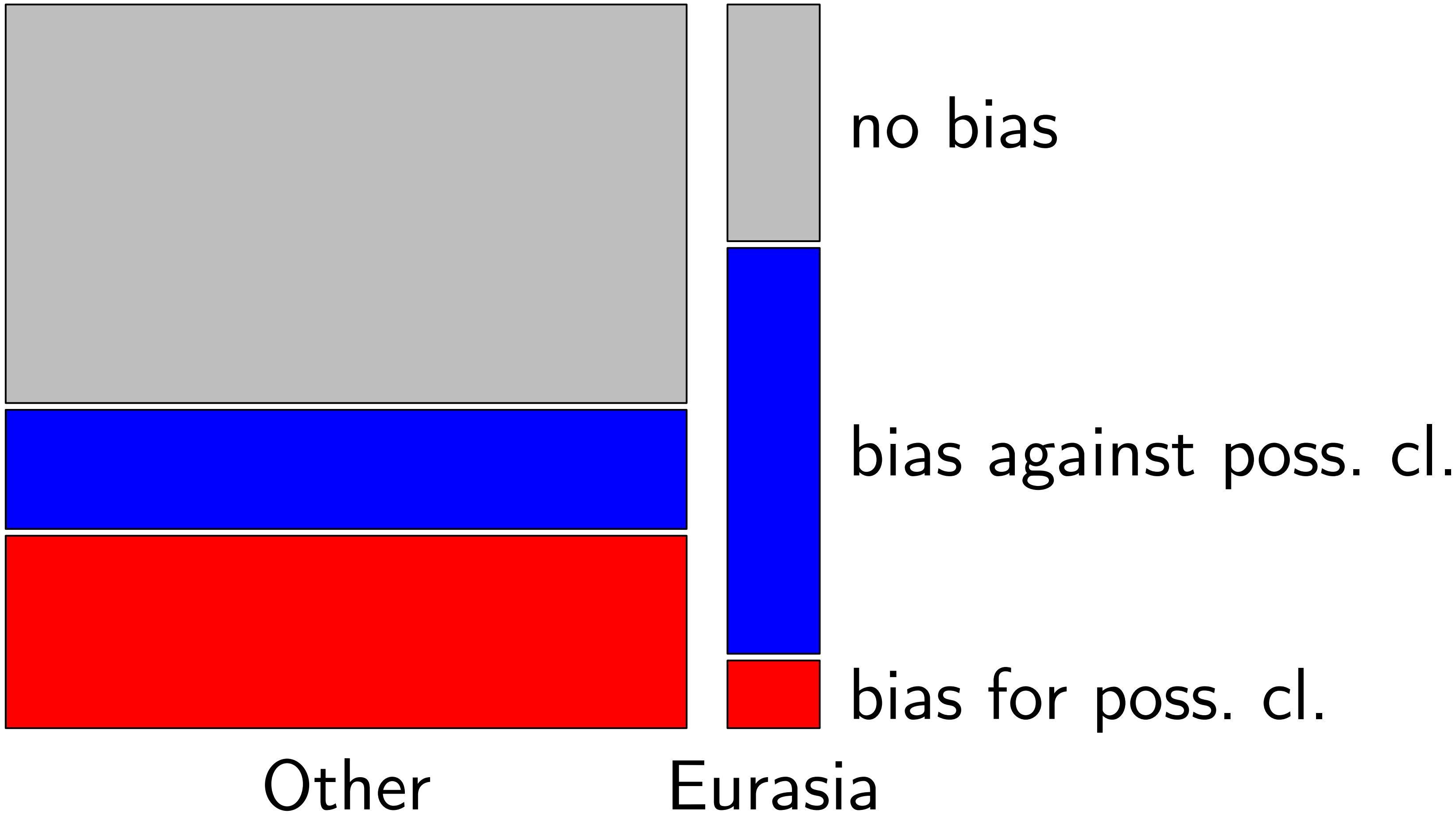
# Extrapolations to small families and isolates



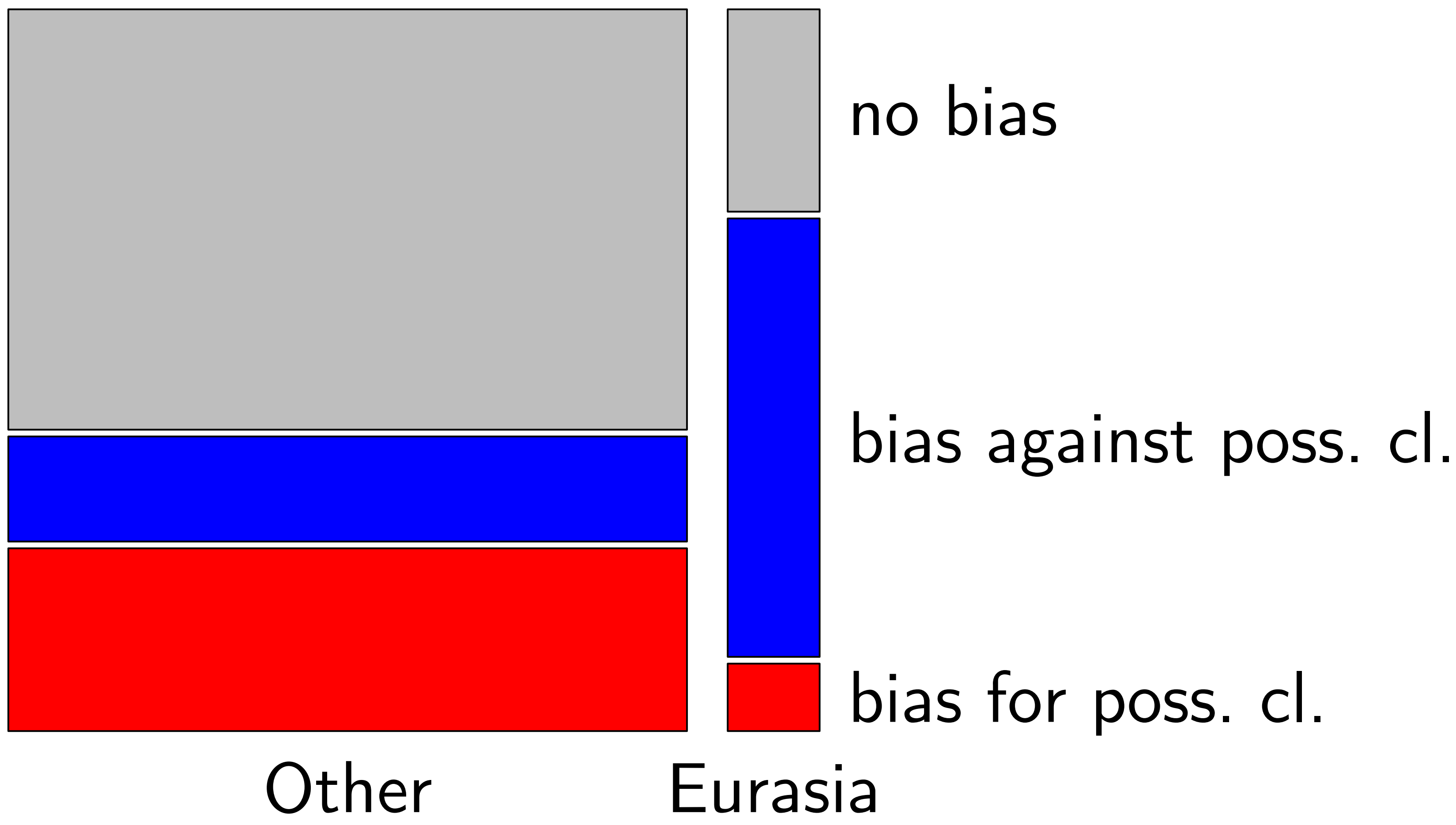
# Extrapolations to small families and isolates



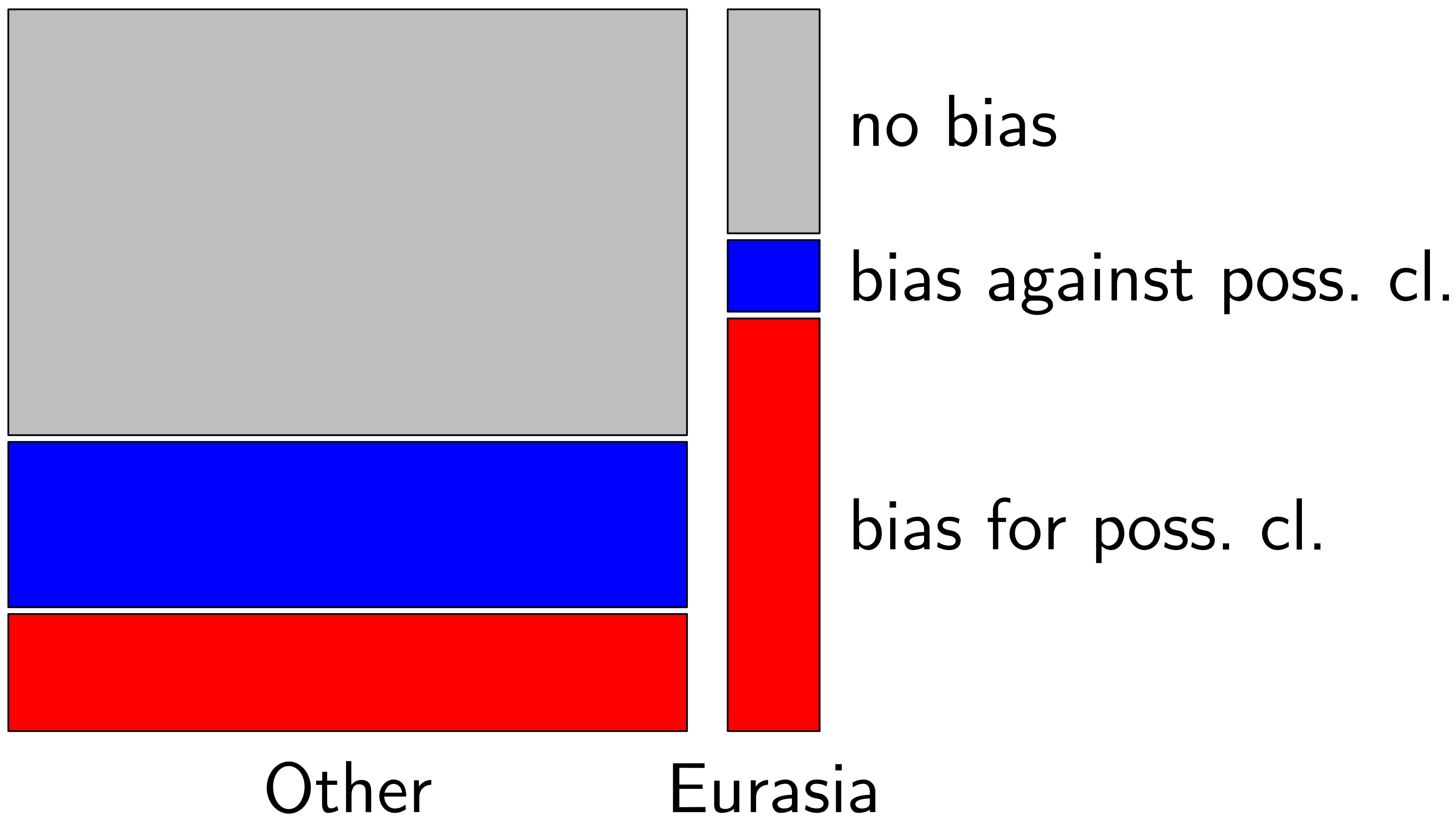
# Extrapolations to small families and isolates



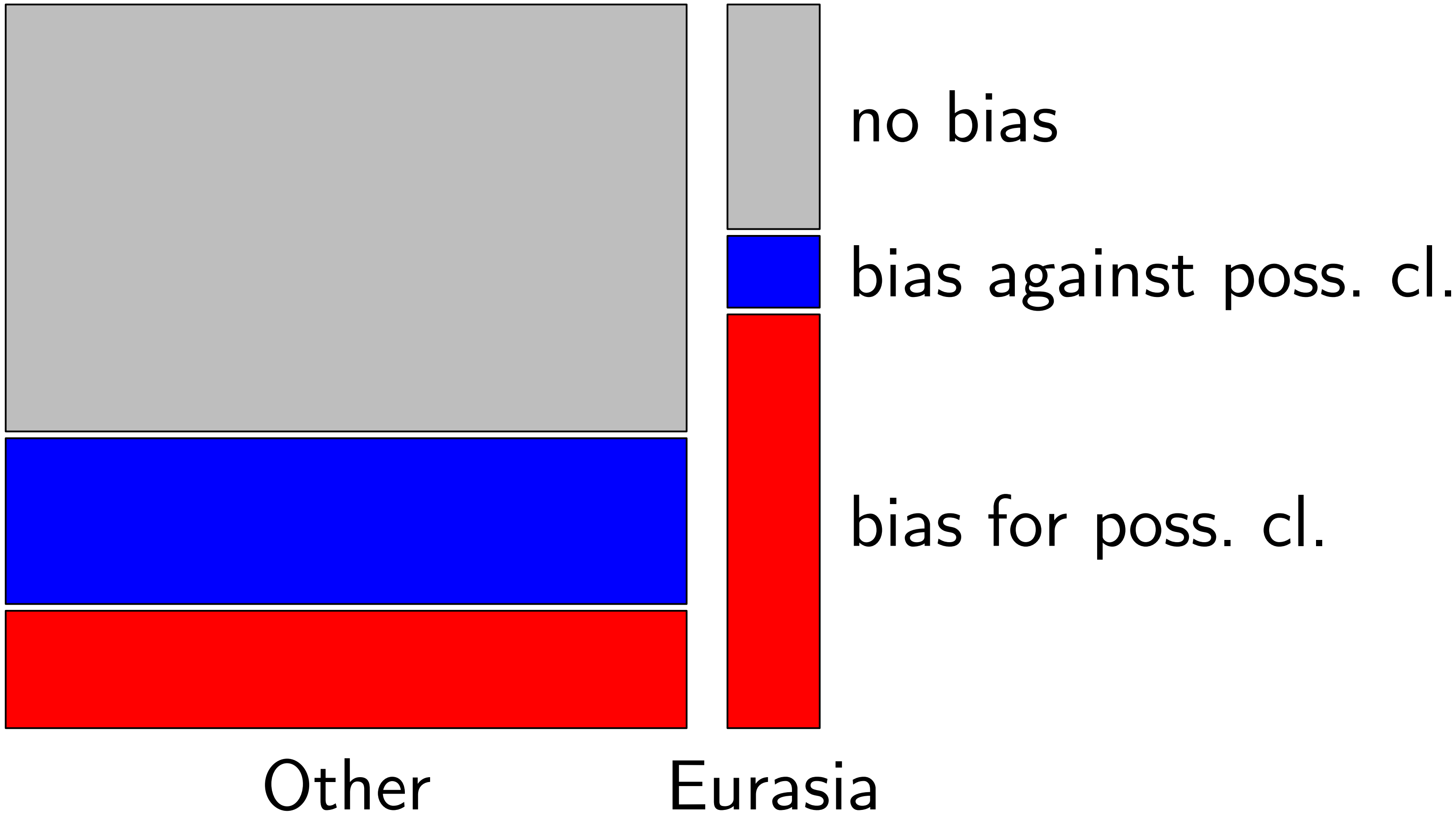
# Extrapolations to small families and isolates



# Extrapolations to small families and isolates



# Extrapolations to small families and isolates




BIAS DIRECTION × AREA: $p = .006$ (Fisher Exact test)

DIVERSITY × AREA: $p = .03$ (Fisher Exact test)

# Evaluating the performance of the Family Bias Method

# Evaluating the performance of the Family Bias Method

- Using the same simulation model as before, same parameters, but now add

# Evaluating the performance of the Family Bias Method

- Using the same simulation model as before, same parameters, but now add
  - Pr(bias): proportion of families with built-in bias vs. absence of a bias in the simulation

# Evaluating the performance of the Family Bias Method

- Using the same simulation model as before, same parameters, but now add
  - Pr(bias): proportion of families with built-in bias vs. absence of a bias in the simulation
  - Pr(bias direction): proportion of families biased towards a specific value (e.g. towards having possessive classes)

# Evaluating the performance of the Family Bias Method

- Using the same simulation model as before, same parameters, but now add
  - Pr(bias): proportion of families with built-in bias vs. absence of a bias in the simulation
  - Pr(bias direction): proportion of families biased towards a specific value (e.g. towards having possessive classes)
- Simulations with various

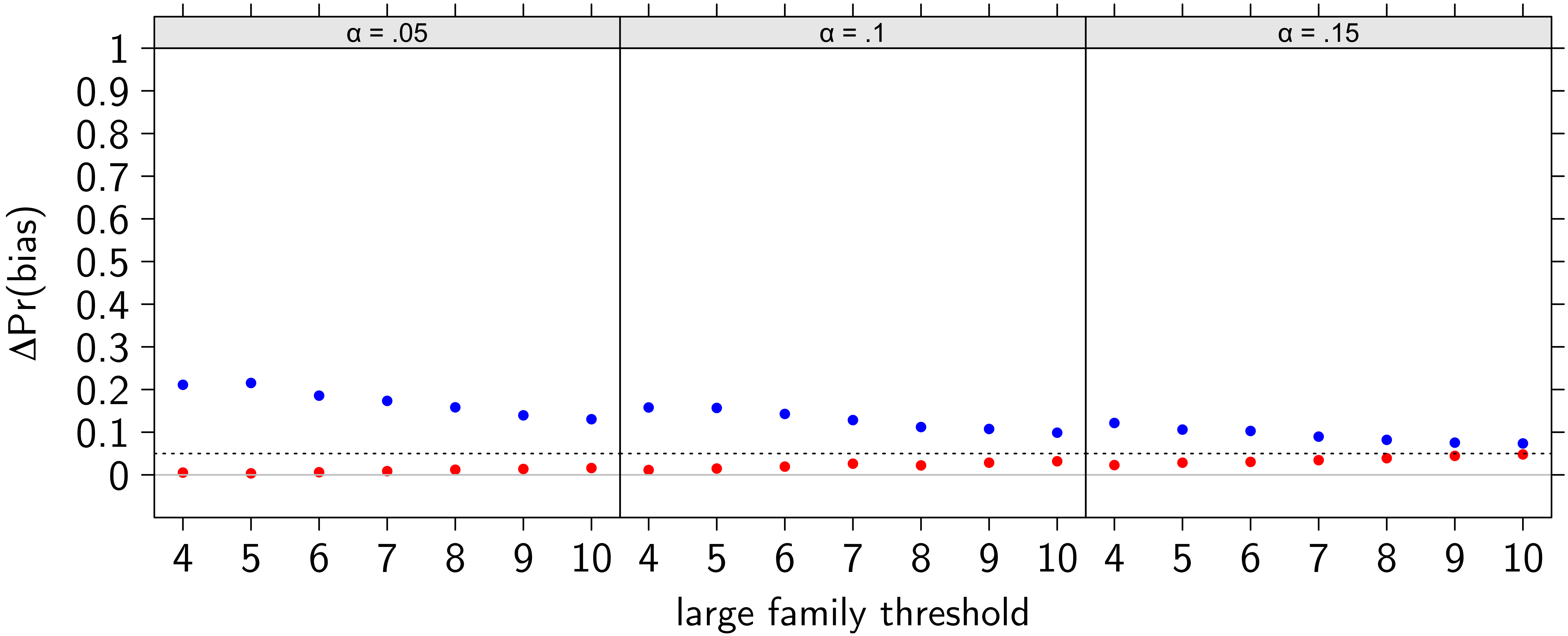
- Using the same simulation model as before, same parameters, but now add
    - Pr(bias): proportion of families with built-in bias vs. absence of a bias in the simulation
    - Pr(bias direction): proportion of families biased towards a specific value (e.g. towards having possessive classes)
- Simulations with various
    - thresholds of what counts as a "large family" vs. what should be left for extrapolation

# Evaluating the performance of the Family Bias Method

- Using the same simulation model as before, same parameters, but now add
  - Pr(bias): proportion of families with built-in bias vs. absence of a bias in the simulation
  - Pr(bias direction): proportion of families biased towards a specific value (e.g. towards having possessive classes)

- Simulations with various
  - thresholds of what counts as a "large family" vs. what should be left for extrapolation
  - rejection levels of the binomial test that evaluates the presence of a biase

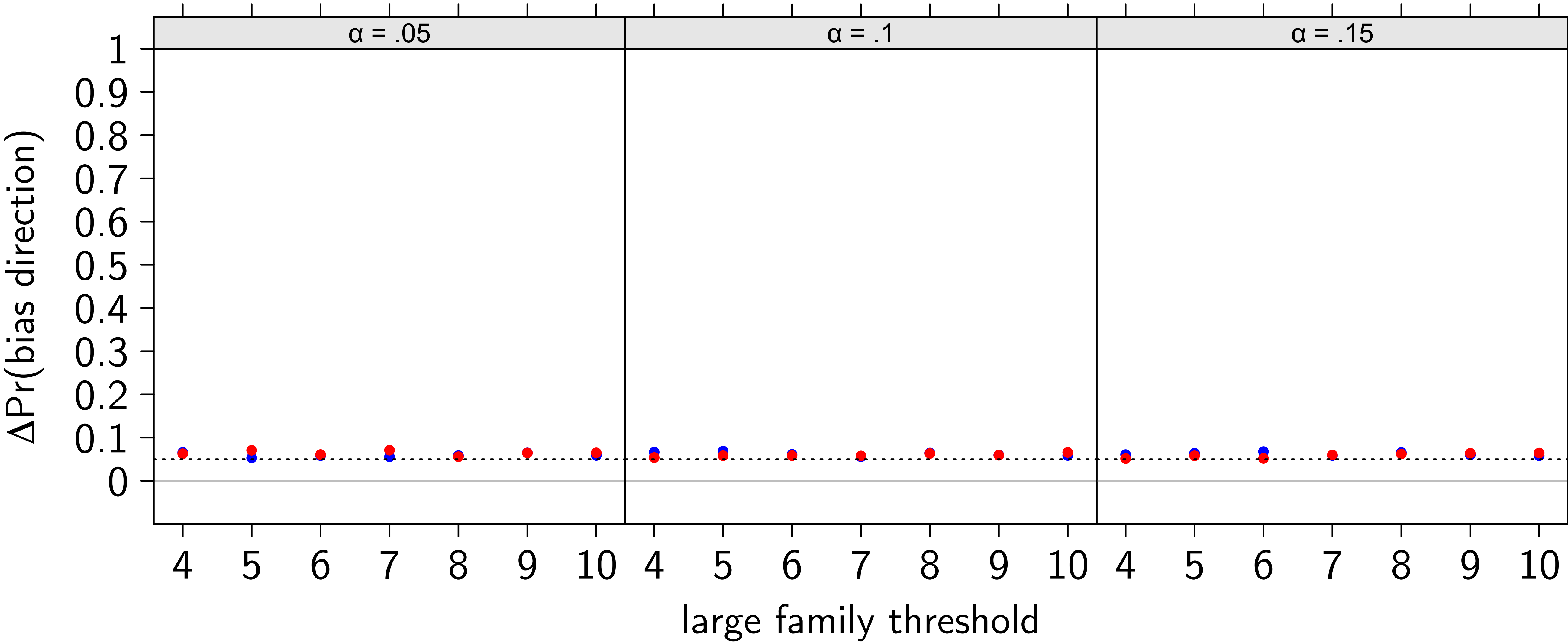
- ΔPr(bias): Absolute difference between Pr(bias) built into the simulation and what is estimated from the results by the Family Bias Method:




Mean Pr(bias) estimated lower than built in

Mean Pr(bias) estimated higher than built in

- $\Delta$Pr(bias direction): Absolute difference between Pr(bias direction) built into the simulation and what is estimated from by the Family Bias Method




Mean Pr(bias direction) estimated lower than built in

Mean Pr(bias direction) estimated higher than built in

# Conclusions

# Conclusions

- The Family Bias Method allows estimation of differences in diachronic trends between areas

## Conclusions

- The Family Bias Method allows estimation of differences in diachronic trends between areas
- Simple binomial tests are conservatively reliable: **overestimation of biases and bias direction $\leq$ .05** (underestimation $\leq$ .21 for biases, $\leq$.07 for bias directions)

# Conclusions

- The Family Bias Method allows estimation of differences in diachronic trends between areas
- Simple binomial tests are conservatively reliable: **overestimation of biases and bias direction ≤ .05** (underestimation ≤ .21 for biases, ≤.07 for bias directions)
    - But binomial tests are just one of many options compatible with the method; any other estimates, e.g. from Bayesian approaches, can be fed into the method as well (work in progress)

# Conclusions

- The Family Bias Method allows estimation of differences in diachronic trends between areas
- Simple binomial tests are conservatively reliable: **overestimation of biases and bias direction ≤ .05** (underestimation ≤ .21 for biases, ≤.07 for bias directions)
  - But binomial tests are just one of many options compatible with the method; any other estimates, e.g. from Bayesian approaches, can be fed into the method as well (work in progress)
- What is essential for the method is that it pays attention

# Conclusions

- The Family Bias Method allows estimation of differences in diachronic trends between areas
- Simple binomial tests are conservatively reliable: **overestimation of biases and bias direction ≤ .05** (underestimation ≤ .21 for biases, ≤.07 for bias directions)
  - But binomial tests are just one of many options compatible with the method; any other estimates, e.g. from Bayesian approaches, can be fed into the method as well (work in progress)
- What is essential for the method is that it pays attention
  - to both differences in innovation **and retention**

## Conclusions

- The Family Bias Method allows estimation of differences in diachronic trends between areas
- Simple binomial tests are conservatively reliable: **overestimation of biases and bias direction $\leq$ .05** (underestimation $\leq$ .21 for biases, $\leq$.07 for bias directions)
  - But binomial tests are just one of many options compatible with the method; any other estimates, e.g. from Bayesian approaches, can be fed into the method as well (work in progress)
- What is essential for the method is that it pays attention
  - to both differences in innovation **and retention**
  - information from large families **and from small families and isolates**