Inferring probabilities of change from areal distributions

Michael Cysouw
Phillips Universität Marburg

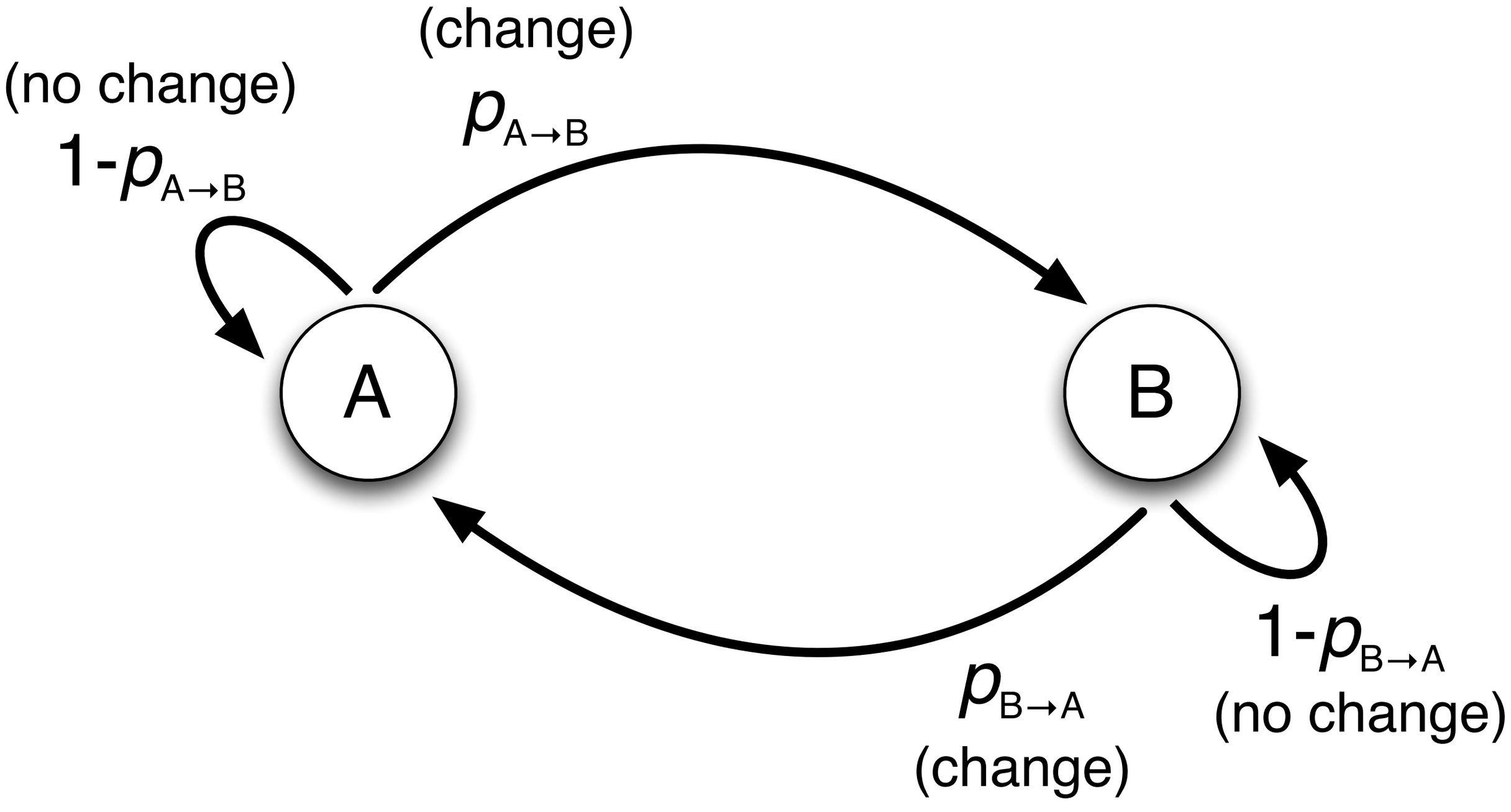
Diachronic and functional explanations in typology

- The explananda are cross-linguistic distributions (aka 'typological patterns' or 'universals')
- My approach: try to restate (not: explain!) typological patterns as diachronic patterns
- Later: try to explain diachronic patterns (this part is still extremely speculative here)

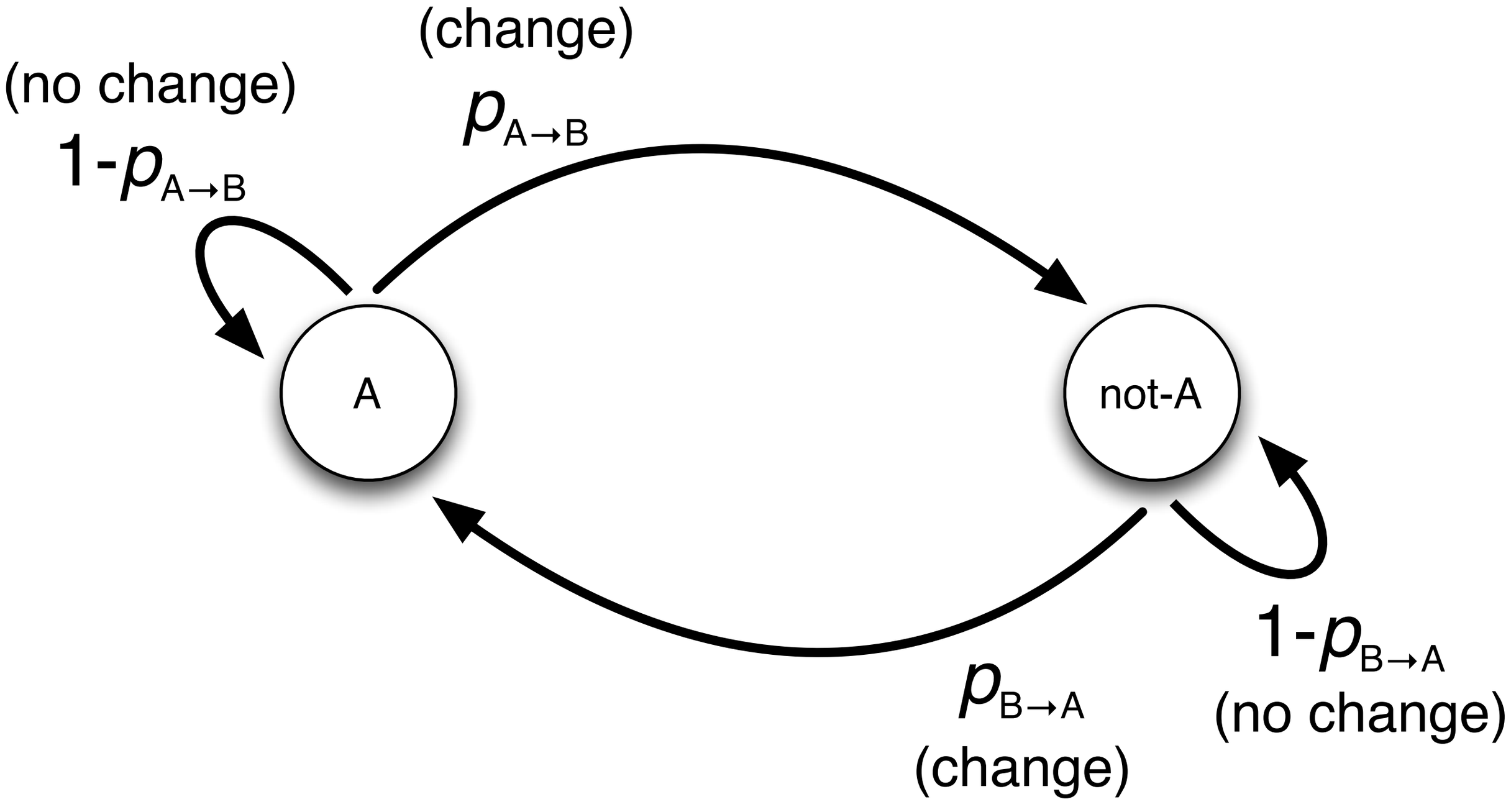
Turn typological patterns into diachronic patterns

- Why?
- Typological patterns/distributions that we can observe in the world are clearly shaped by historical coincidences (so: actual≠possible)
- Deriving diachronic patterns from observed data is an attempt to abstract away from the coincidences skewing typological frequencies

Probabilistic reformulation of change



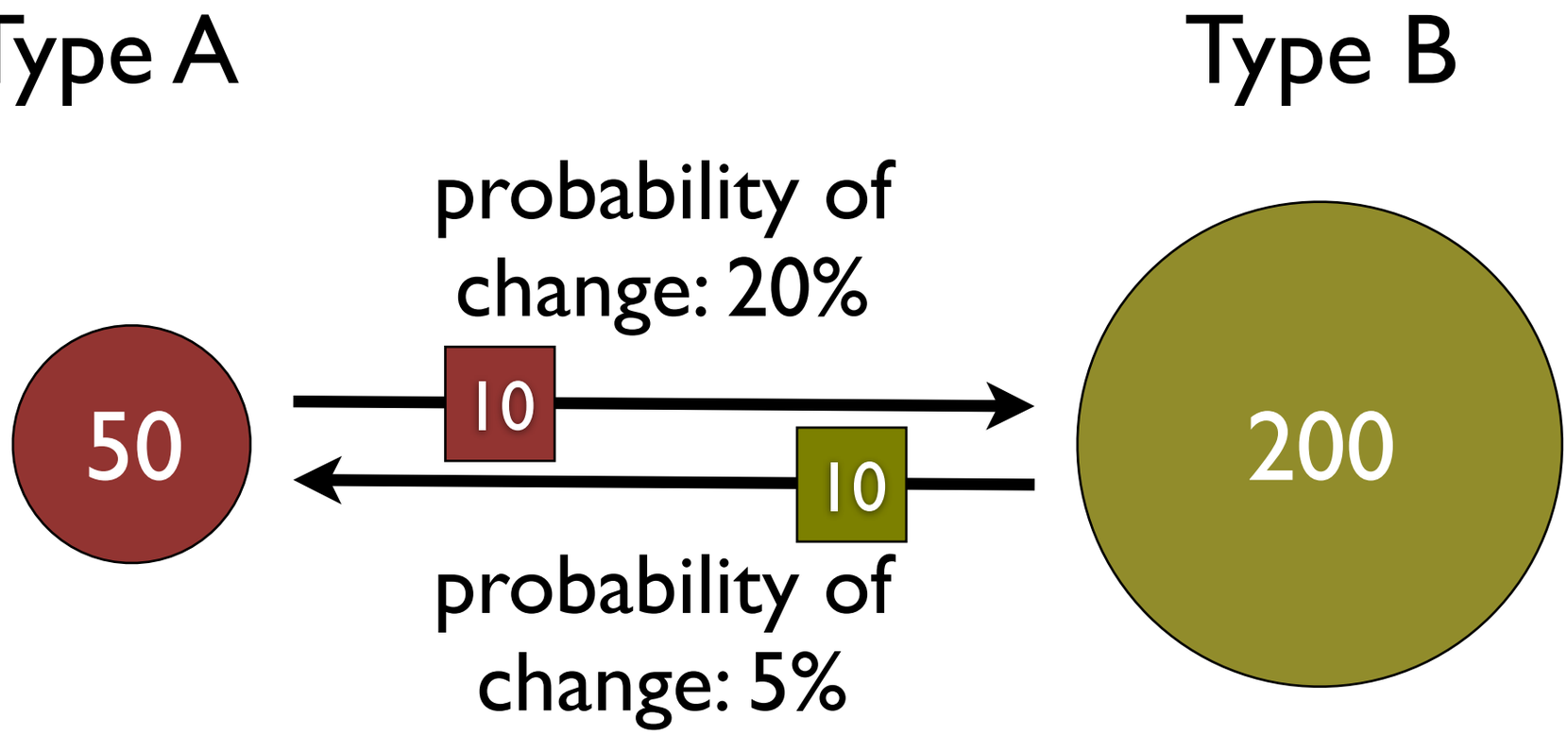
Probabilistic reformulation of change



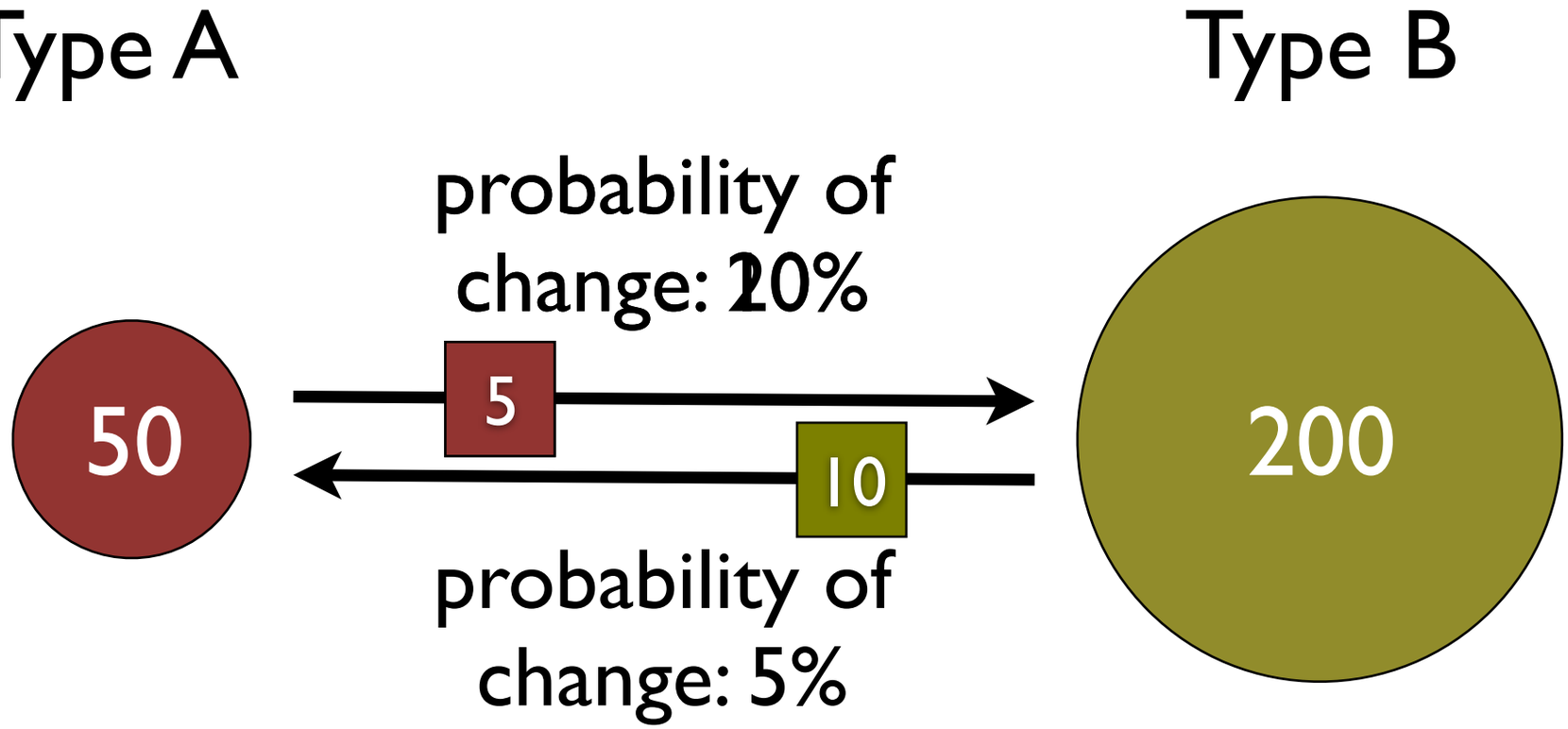
Dynamic Typology

(Maslova 2002, Dediu & Cysouw 2013, cf. Dunn et al. 2011)

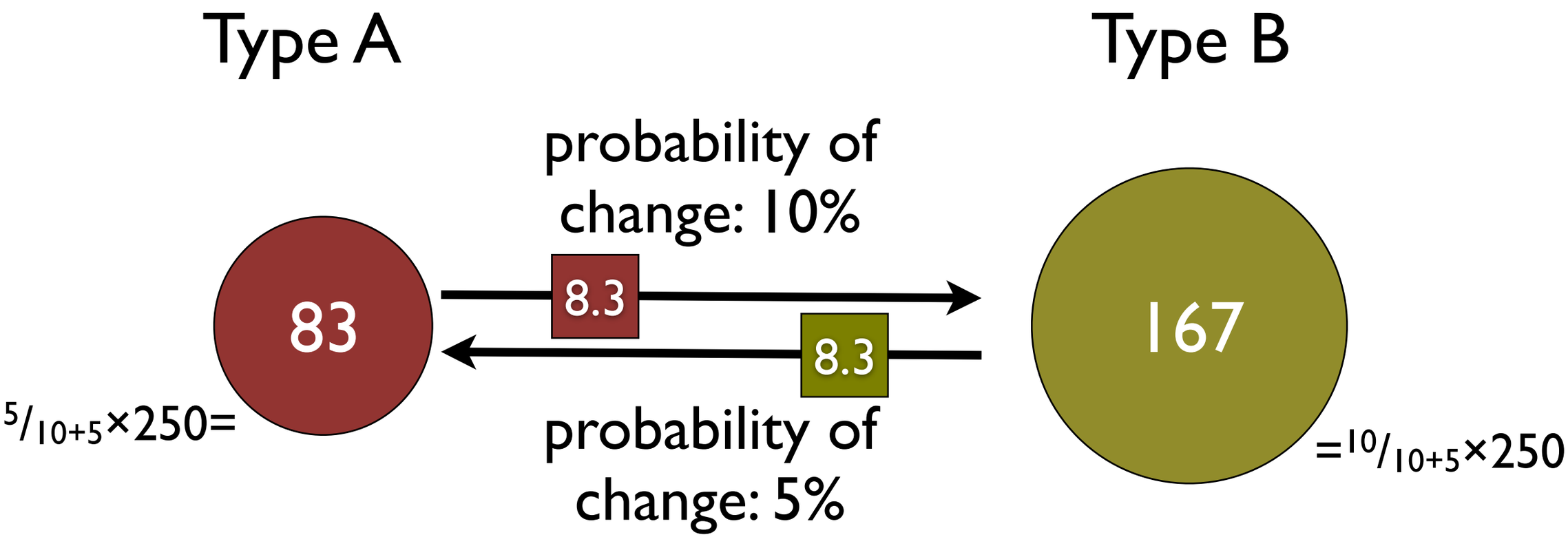
- It is not the **actual frequencies** that matter
- It is the **stable distribution** that matters
  - a stable distribution is a situation in which just as many languages change from A to B as change from B to A.
  - The extent to which the actual is different from the stable situation signals an effect of history



Stable distribution



Instable distribution



Expected stable distribution

Estimating Transition Probabilities

- Are transitions probabilities measurable at all ?
- Use group internal variation of many groups !
- For example:
  - Instead of 100 genealogically unrelated languages
  - take 25 groups of 4 closely related languages
  - Ideally, take related languages with knowledge about the historical relationship (cf. Dunn et al. 2011)

Dunn et al. 2011

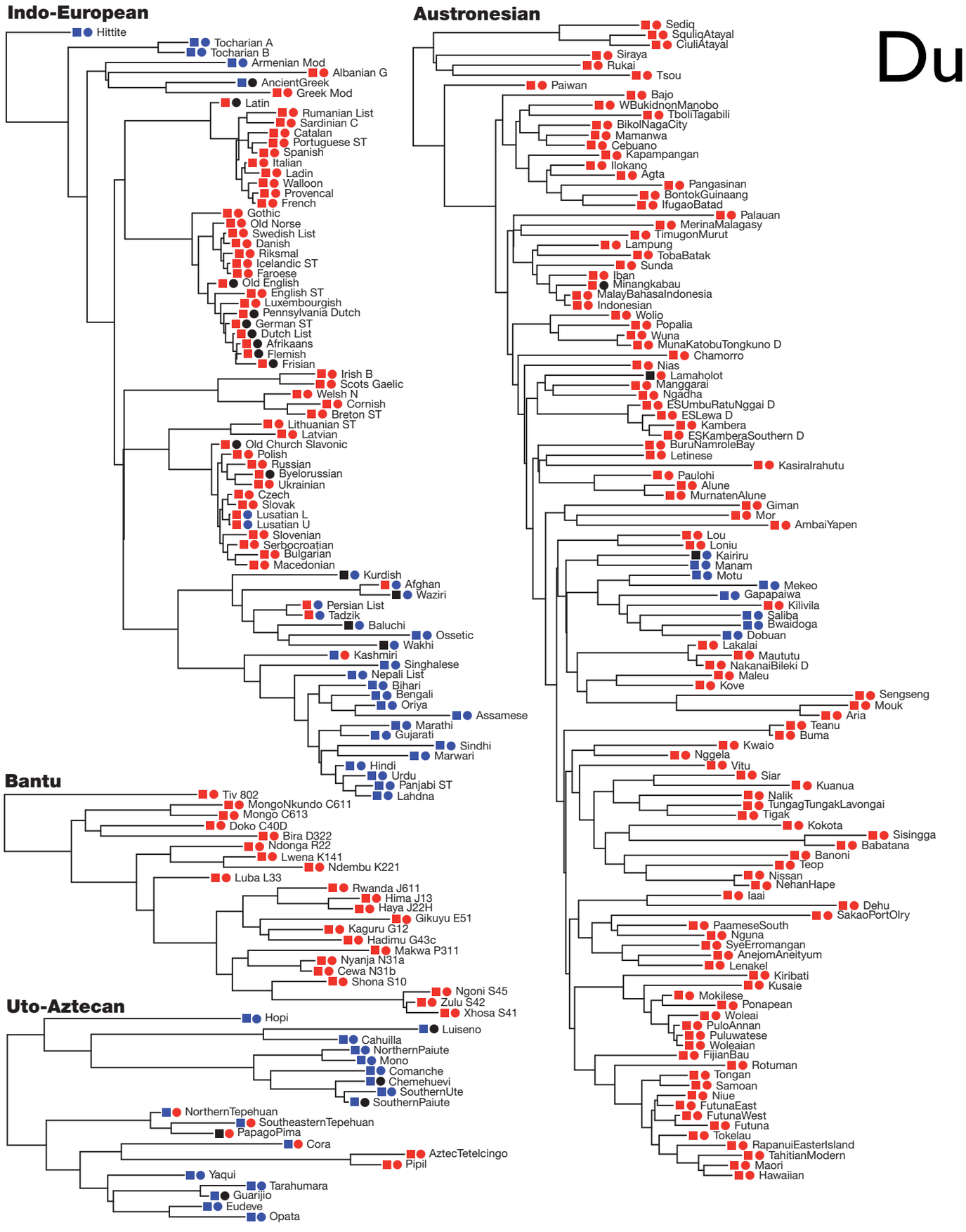
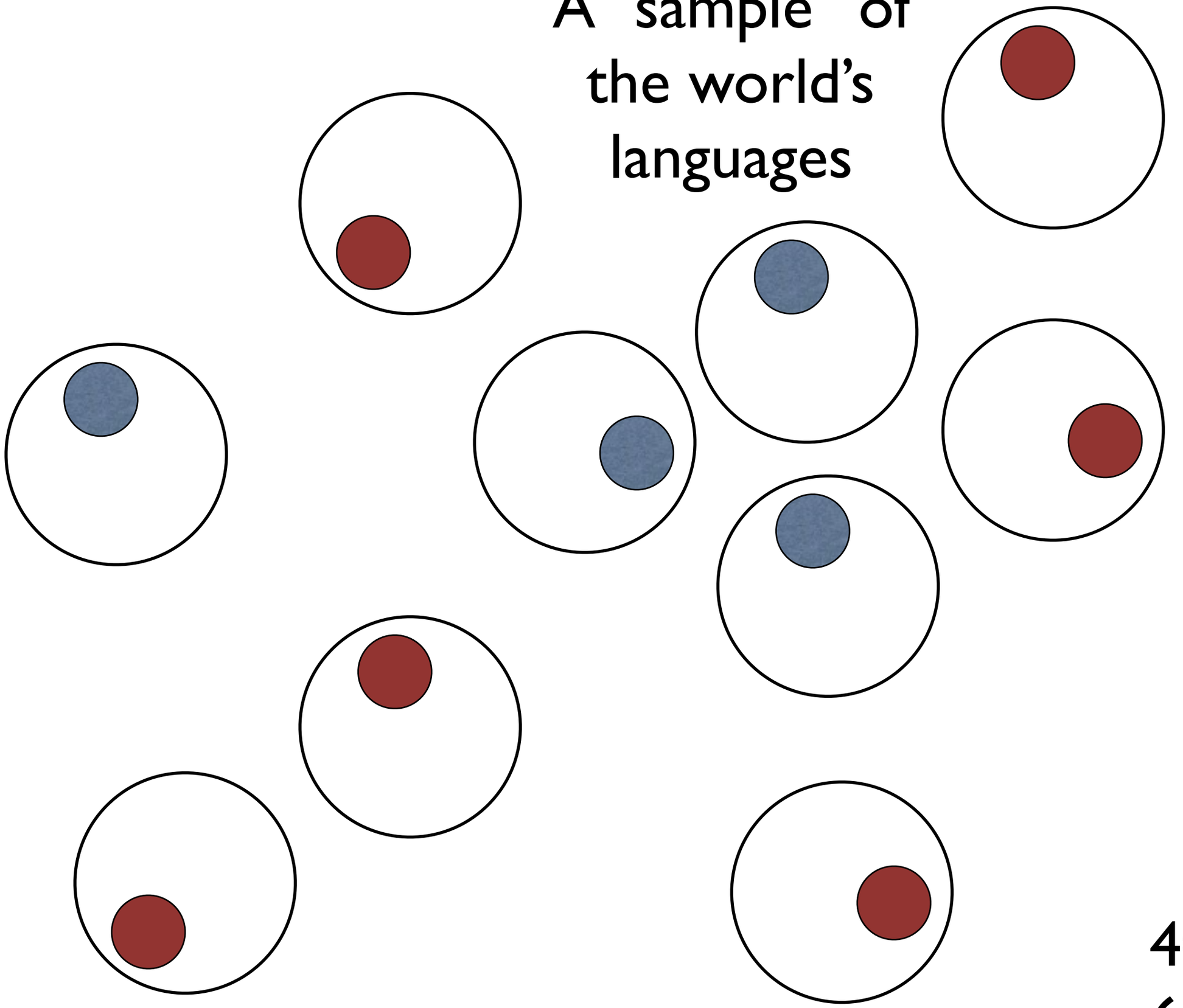


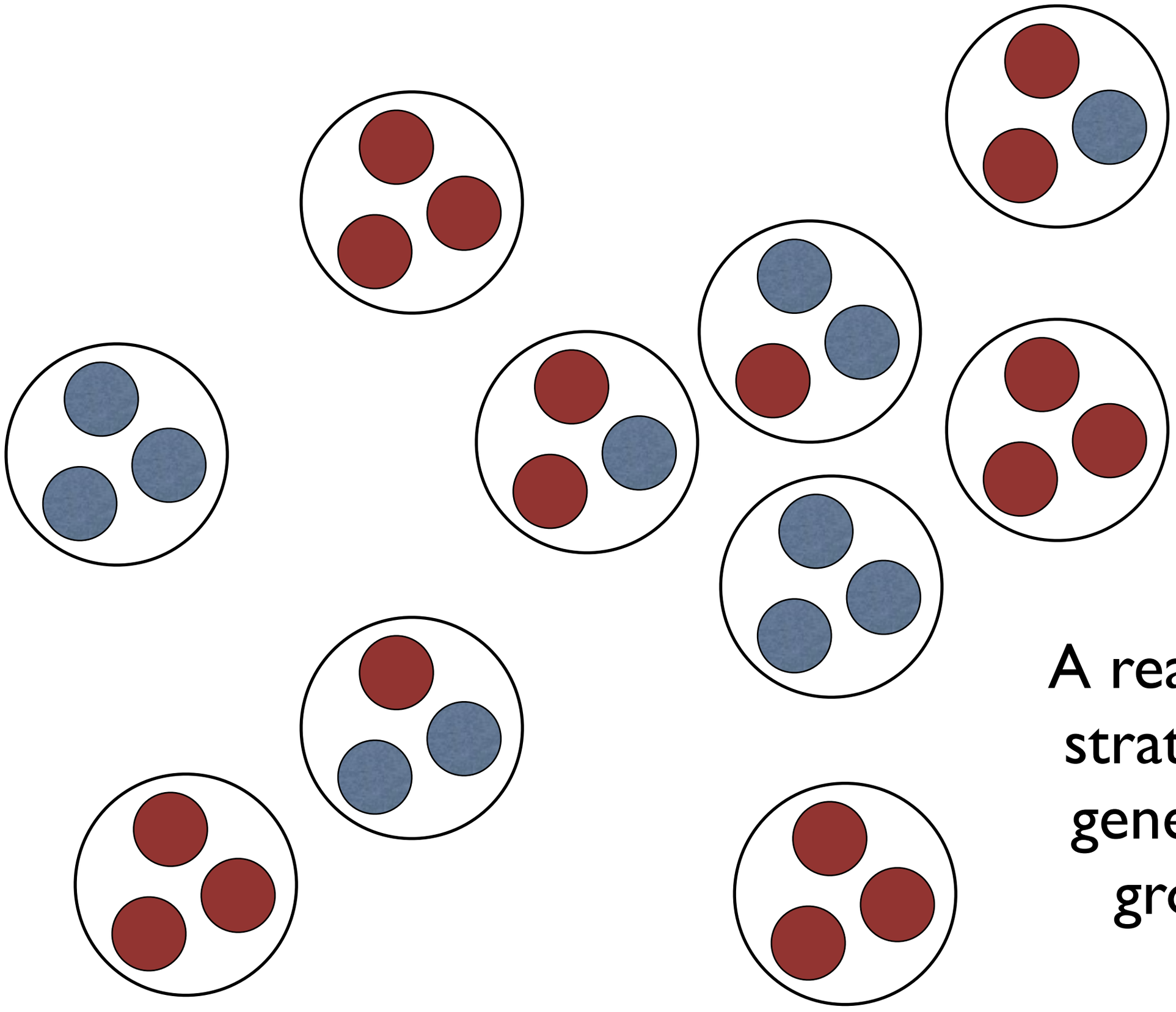
Figure 1 | Two word-order features plotted onto maximum clade credibility trees of the four language families. Squares represent order of adposition and noun; circles represent order of verb and object. The tree sample underlying this tree is generated from lexical data[16,22]. Blue-blue indicates postposition, object–verb. Red-red indicates preposition, verb–object. Red-blue indicates preposition, object–verb. Blue-red indicates postposition, verb–object. Black indicates polymorphic states.

The world's genera

A "sample" of the world's languages

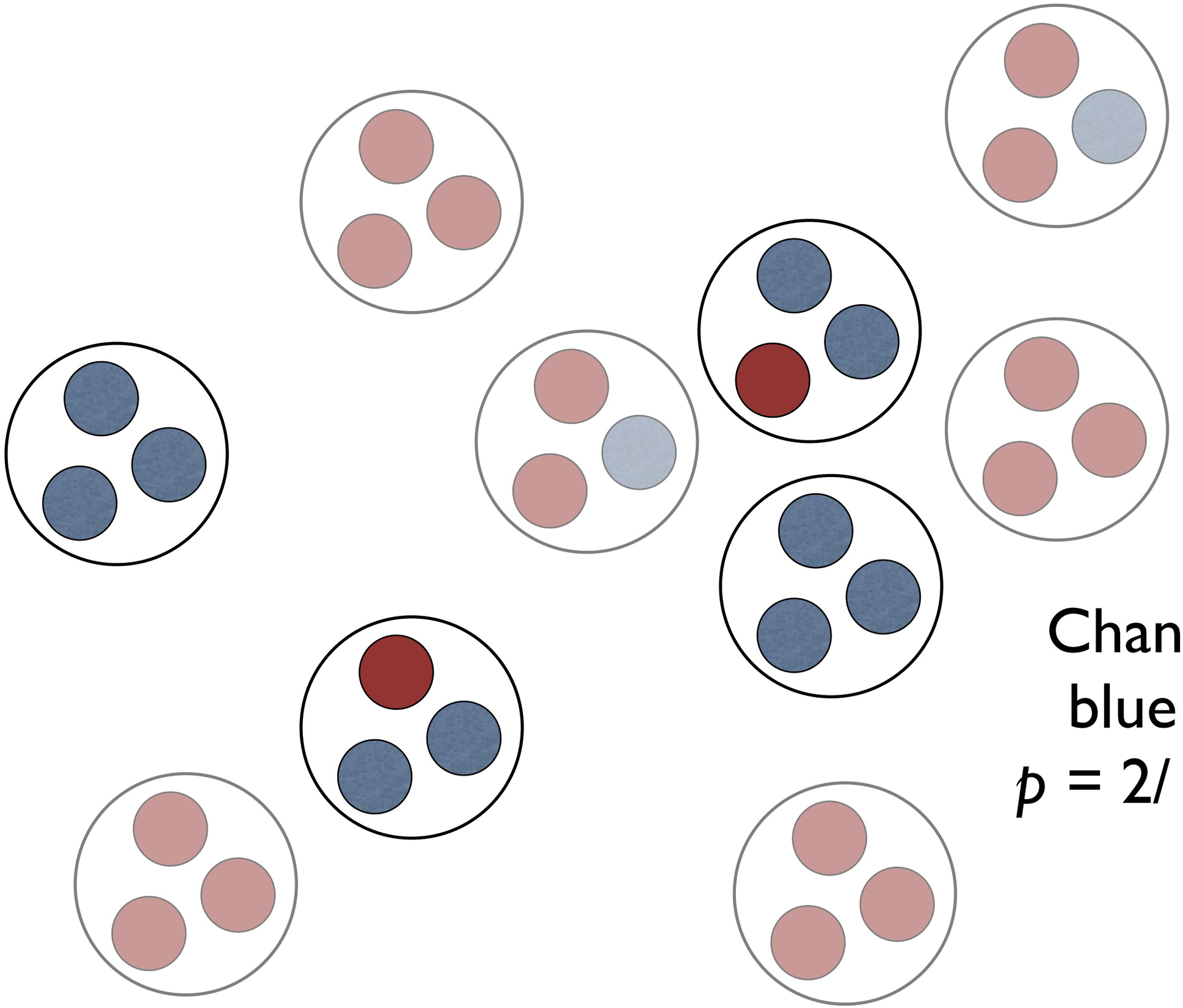


40% blue
60% red

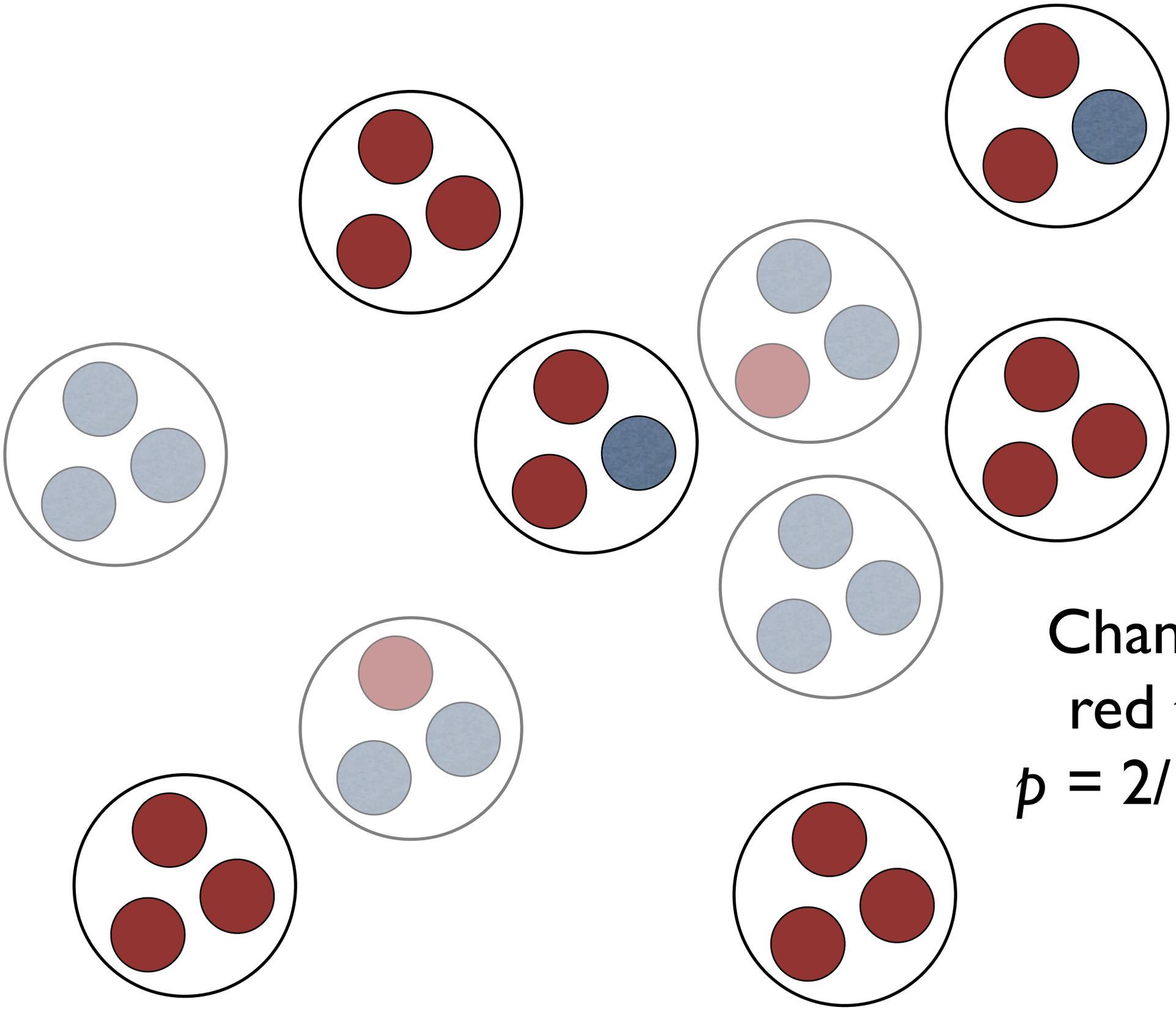


A real sample stratified for genealogical grouping

How to get probabilities of change ...

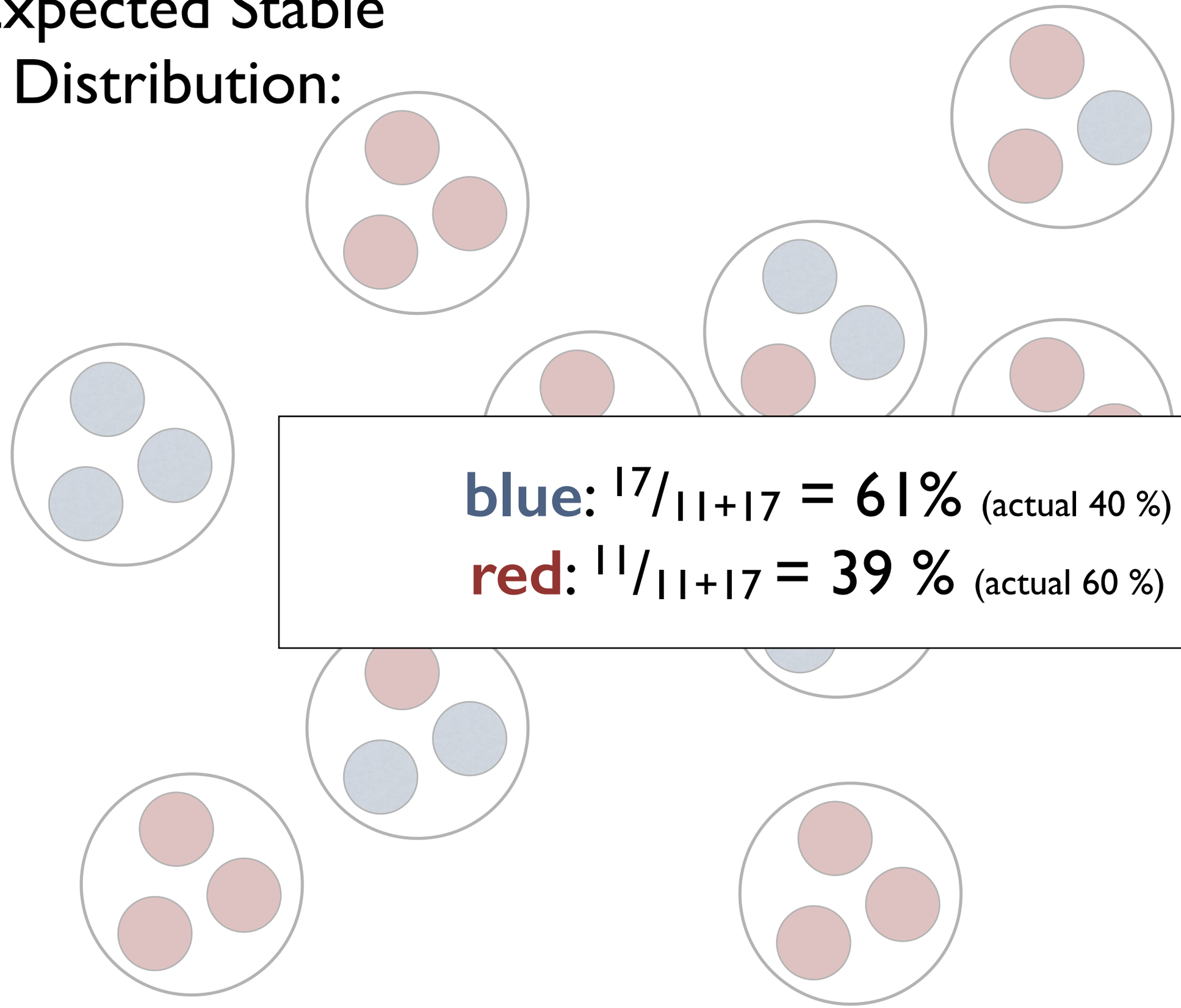


Change from blue to red:
p = 2/12 = 17%



Change from red to blue:
p = 2/18 = 11%

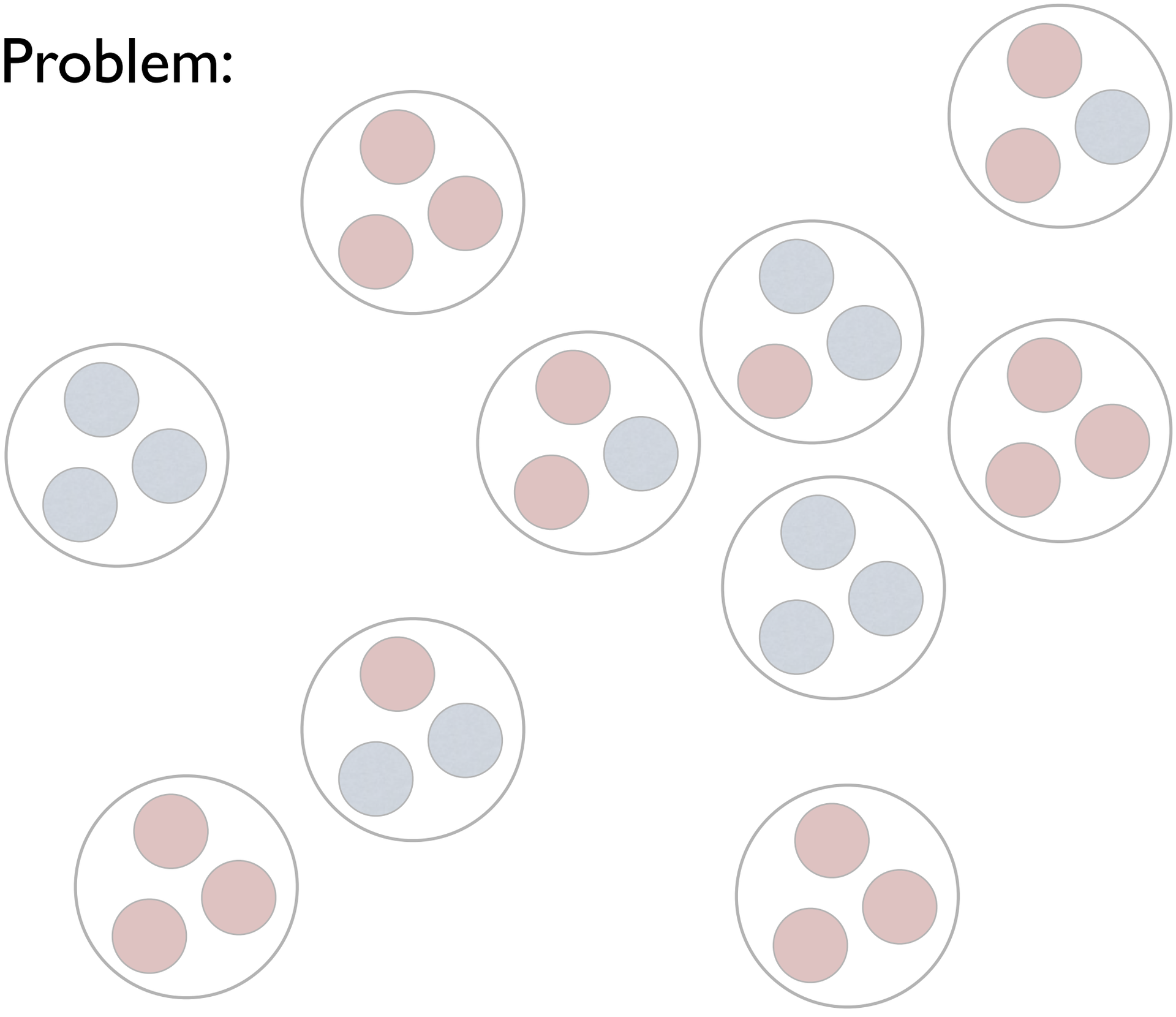
Expected Stable Distribution:



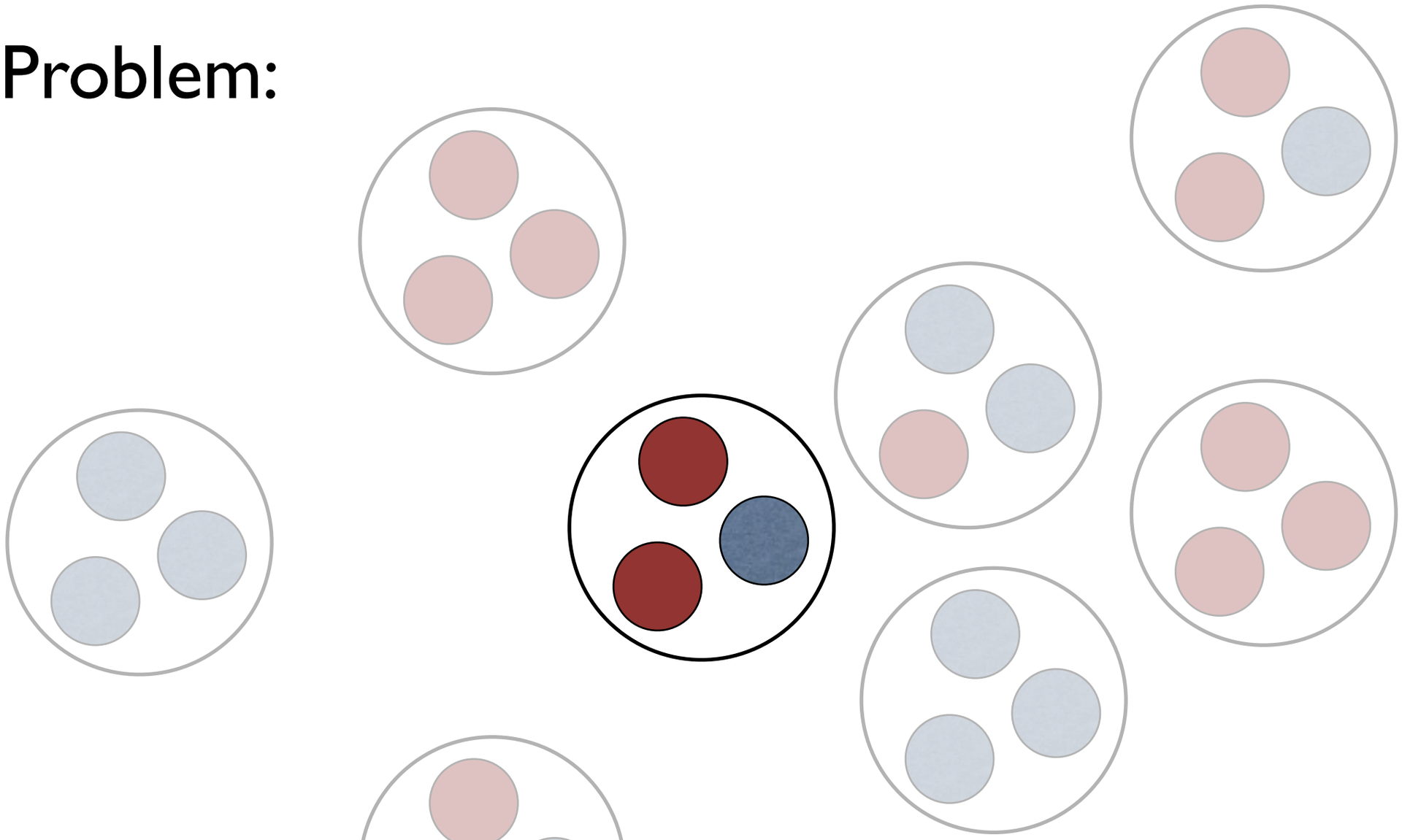
blue: $^{17}/_{11+17} = 61\%$ (actual 40 %)

red: $^{11}/_{11+17} = 39\ \%$ (actual 60 %)

Problem:

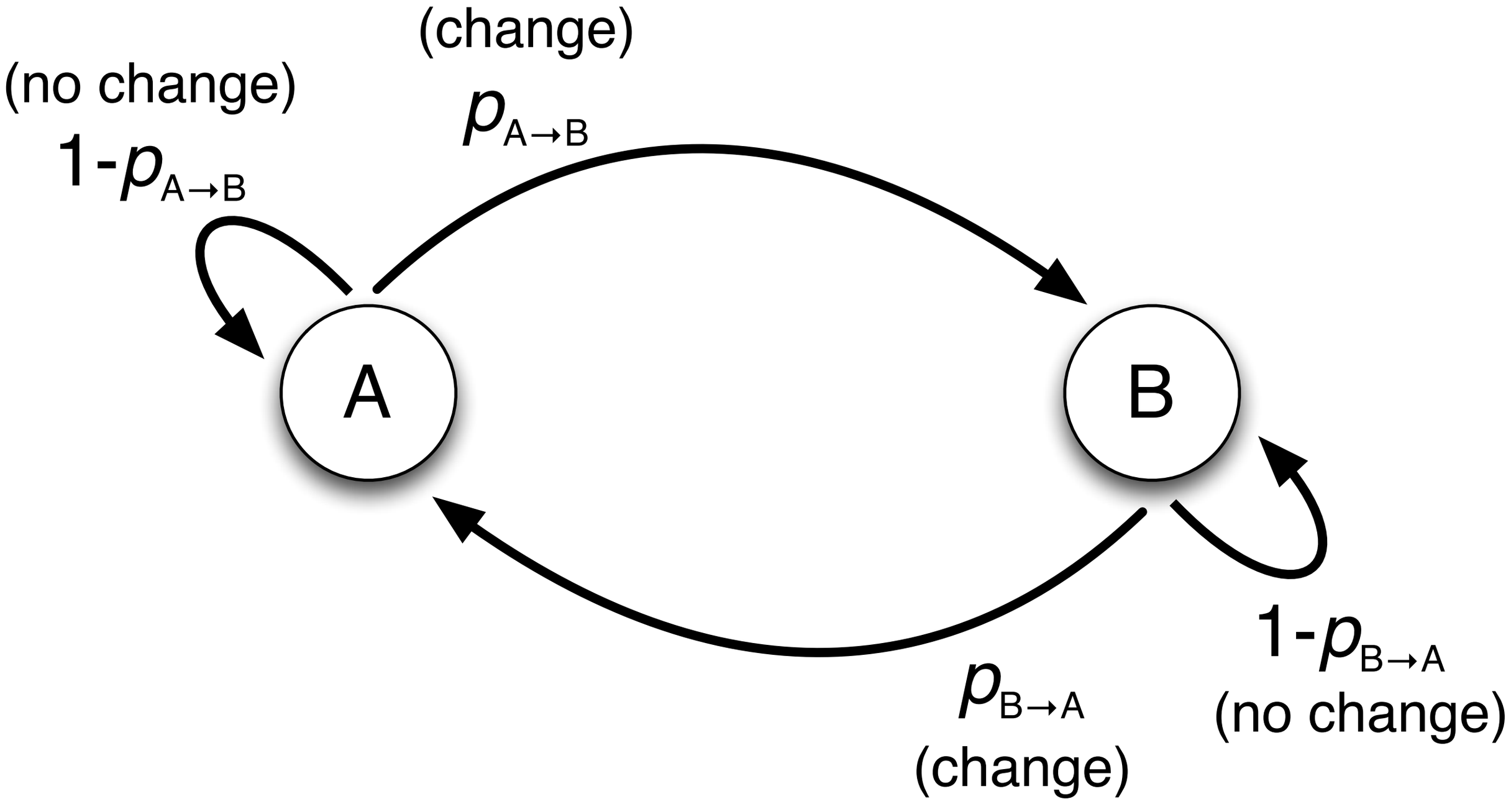


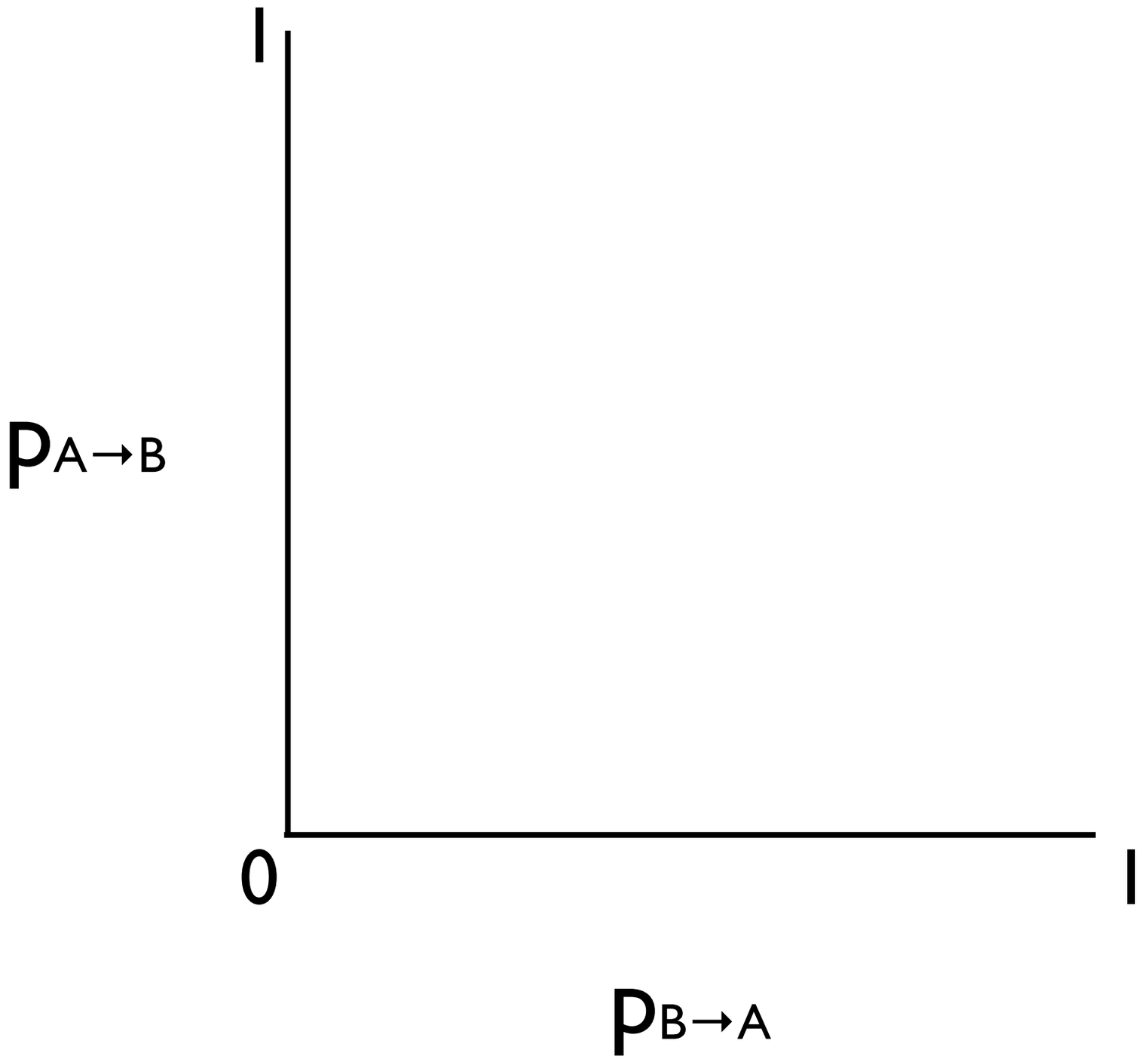
Problem:

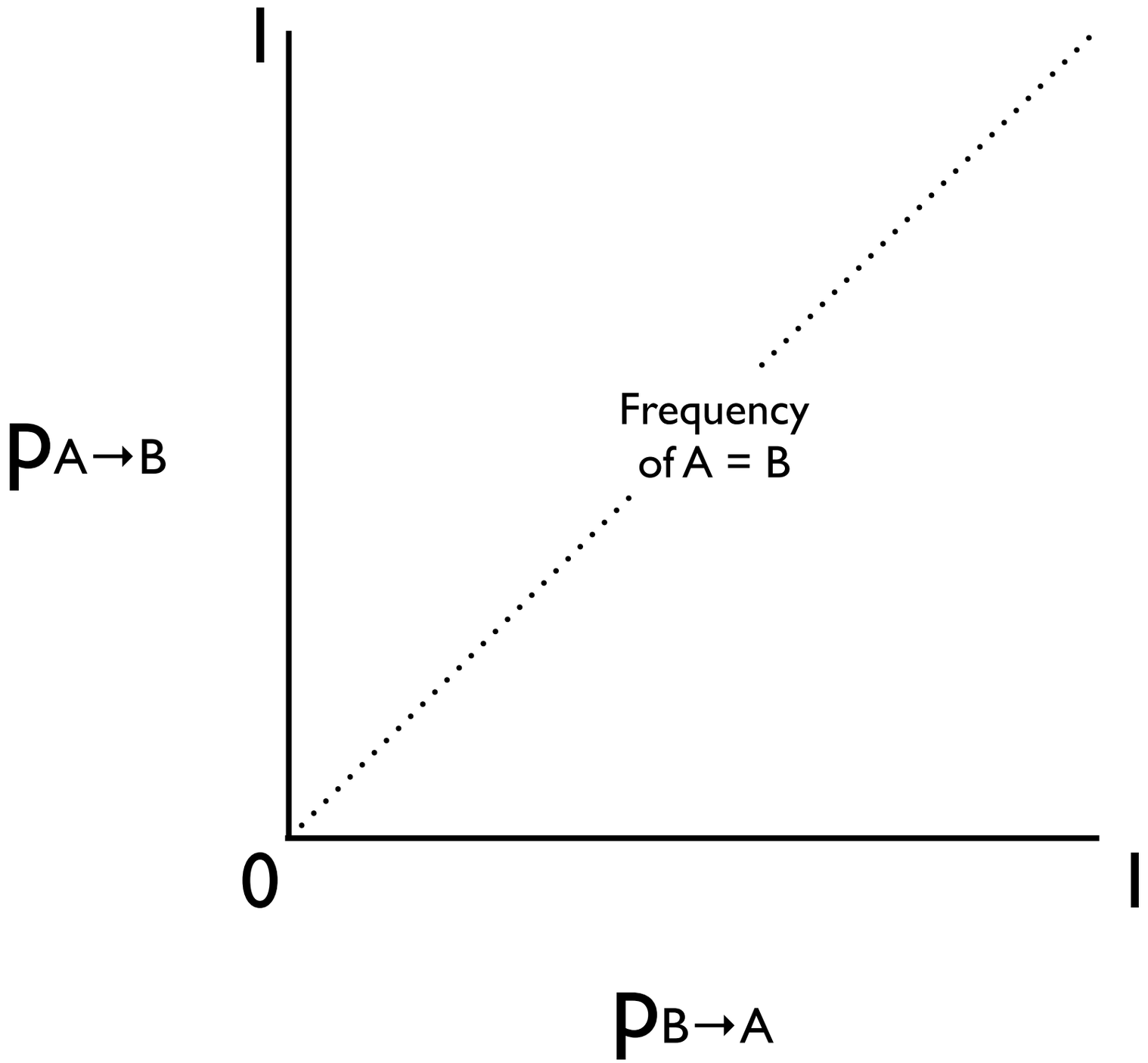


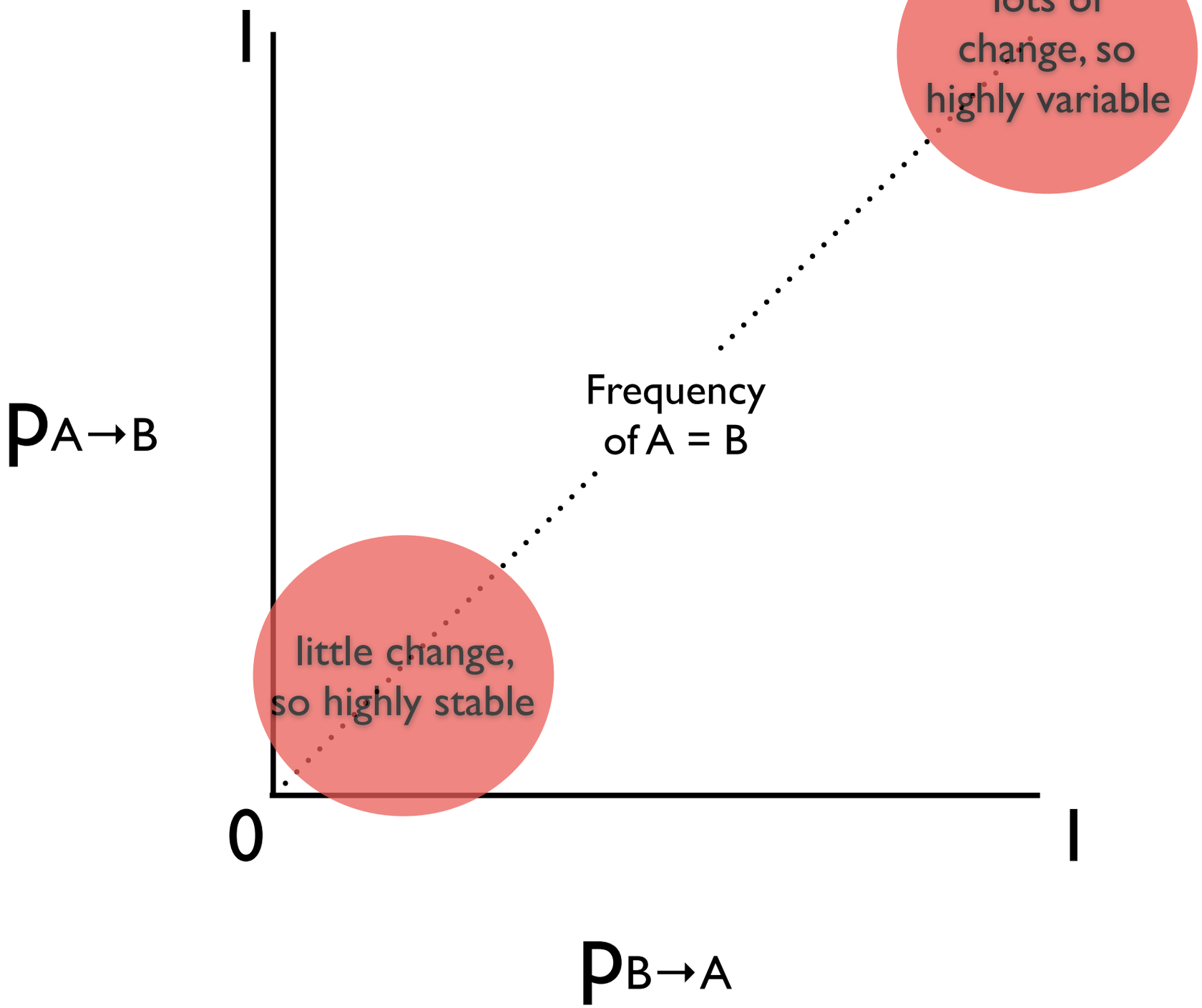
- Originally **red** with one change to blue ?
- Originally **blue** with two changes to red ?

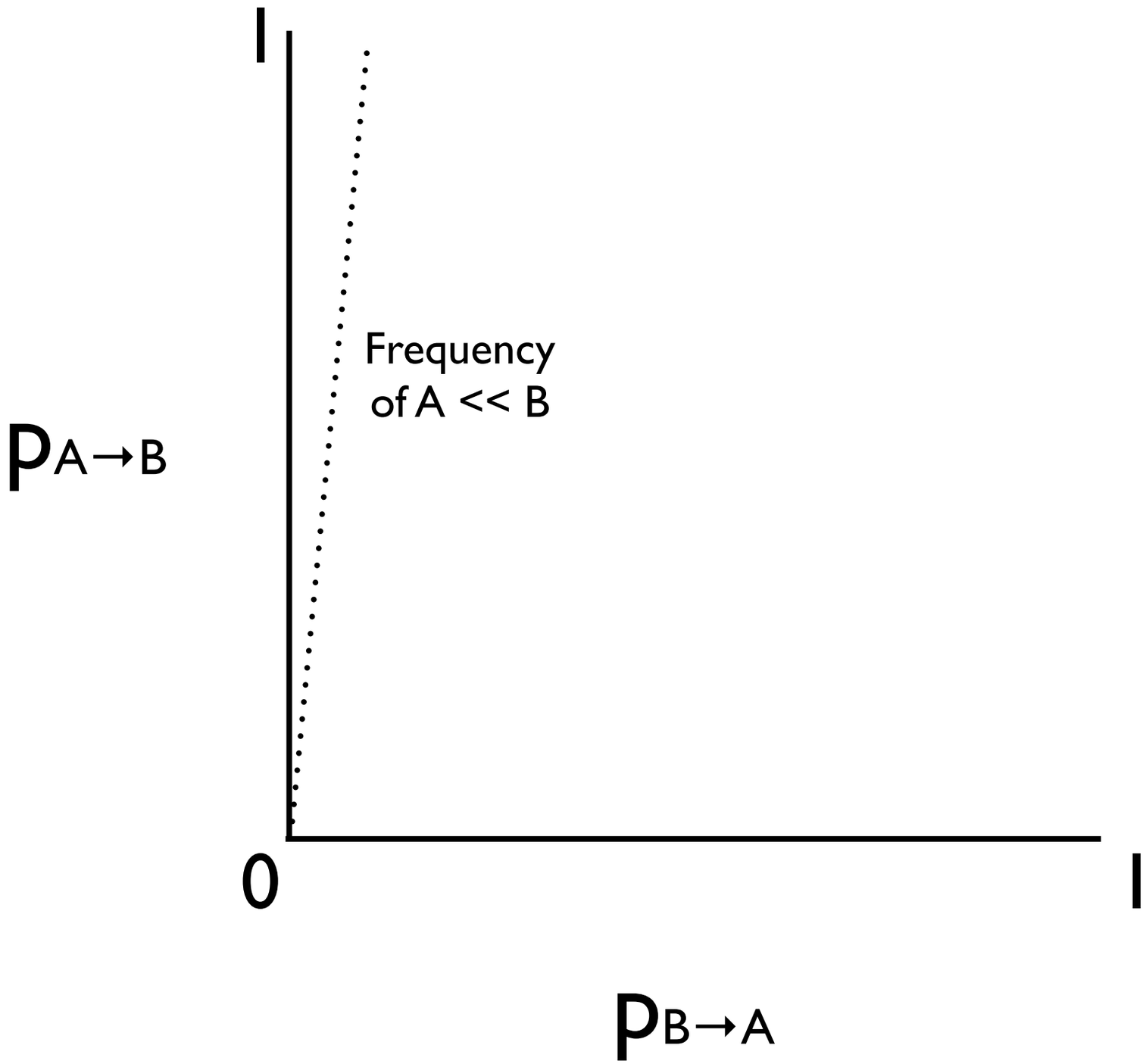
Probabilistic reformulation of change

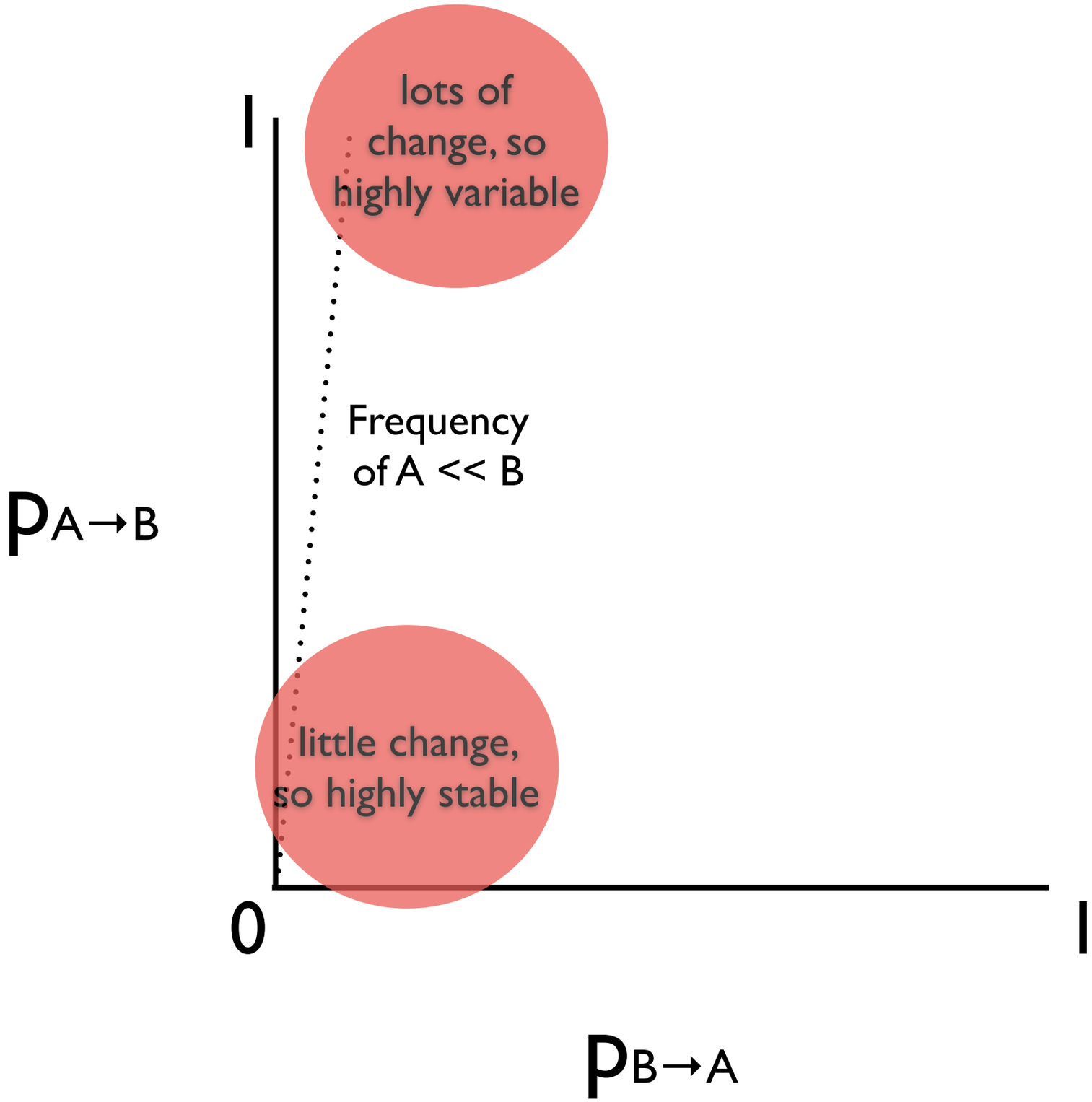



1
$p_{A\to B}$
0
1
$p_{B\to A}$


1
$p_{A \to B}$
Frequency
of A = B
0
1
$p_{B \to A}$

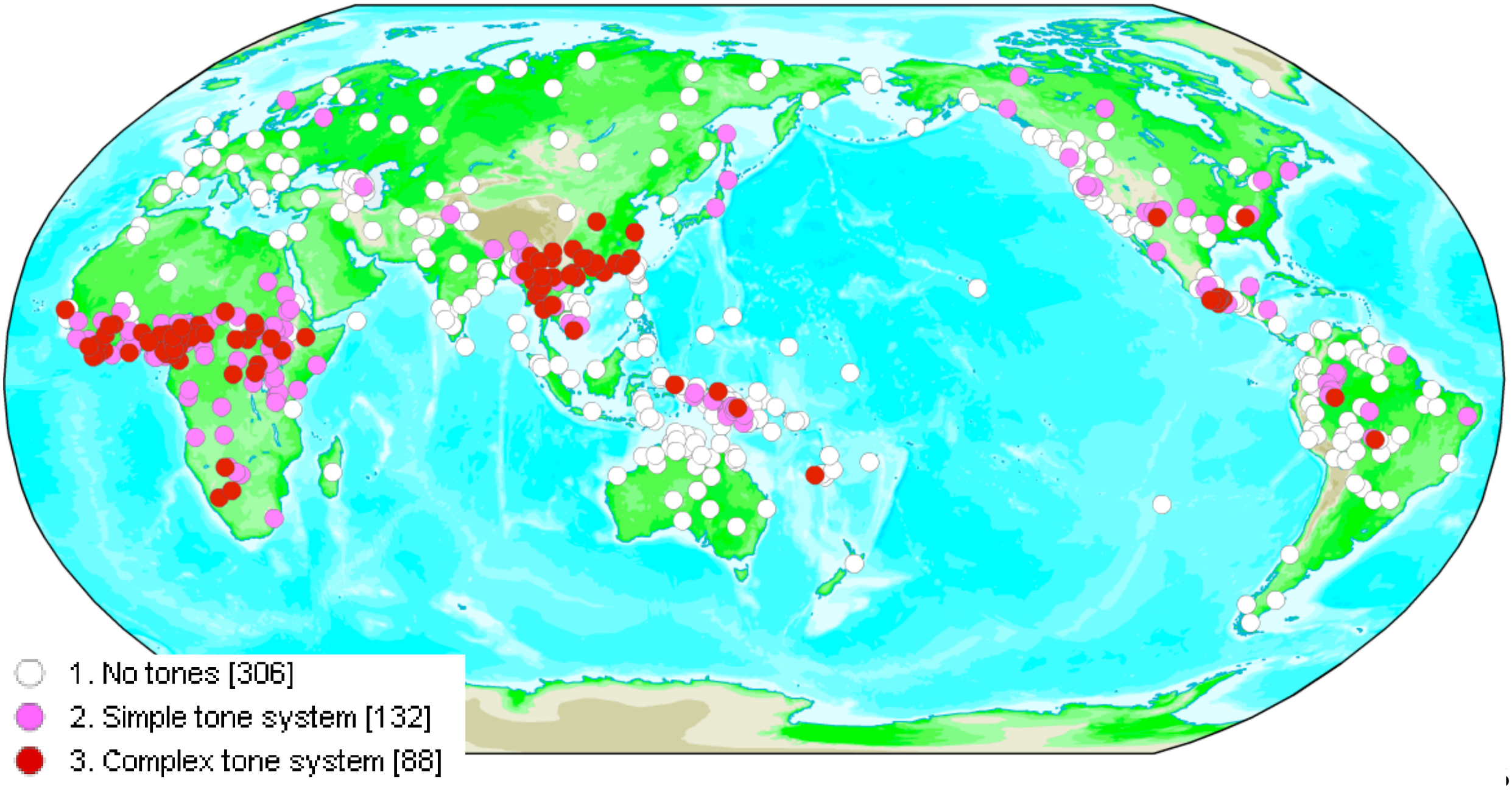

lots of change, so highly variable
Frequency of A = B
little change, so highly stable
1
$p_{A \to B}$
0
1
$p_{B \to A}$

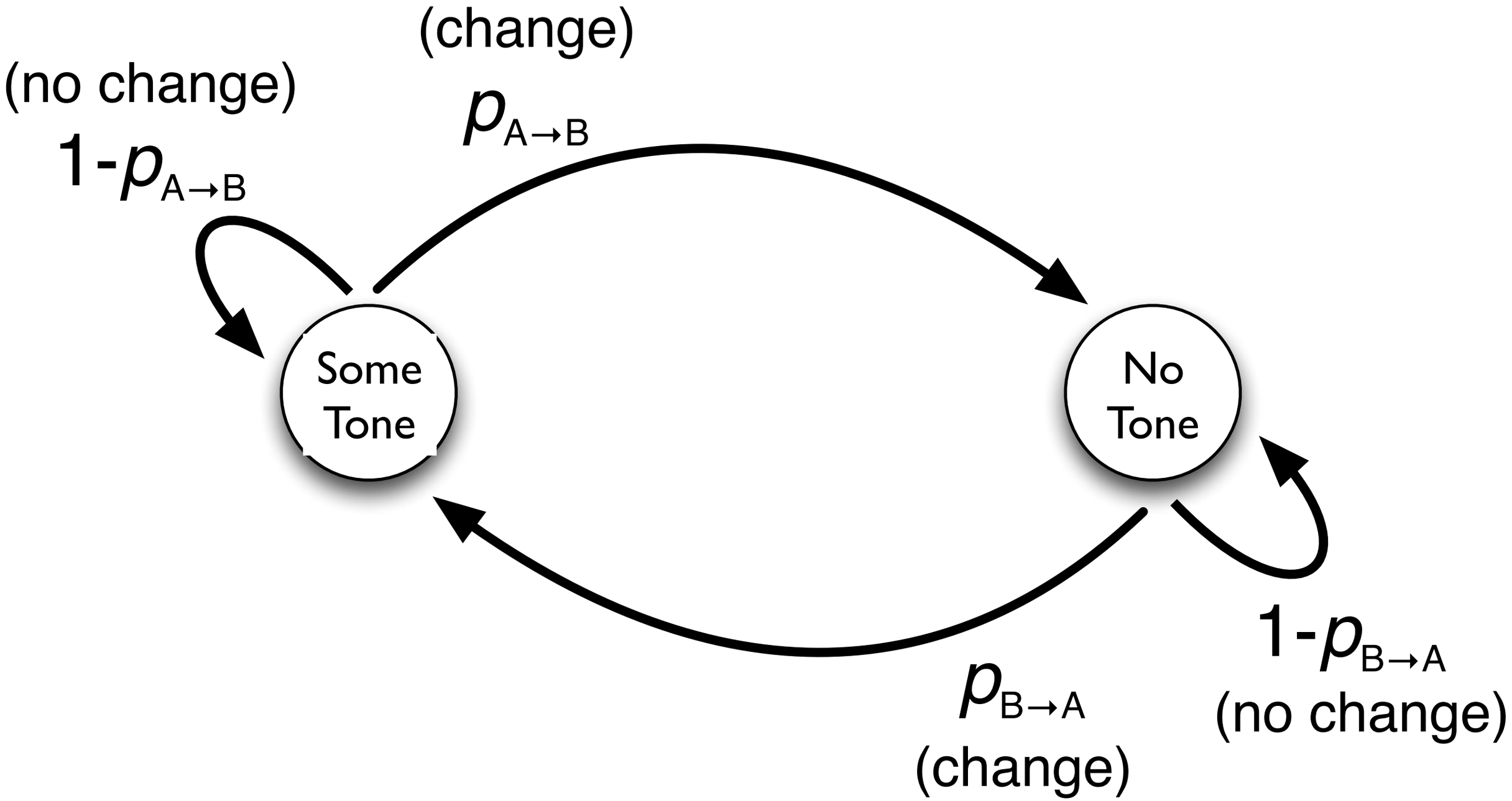

1
Frequency
of A << B
pA→B
0
1
pB→A

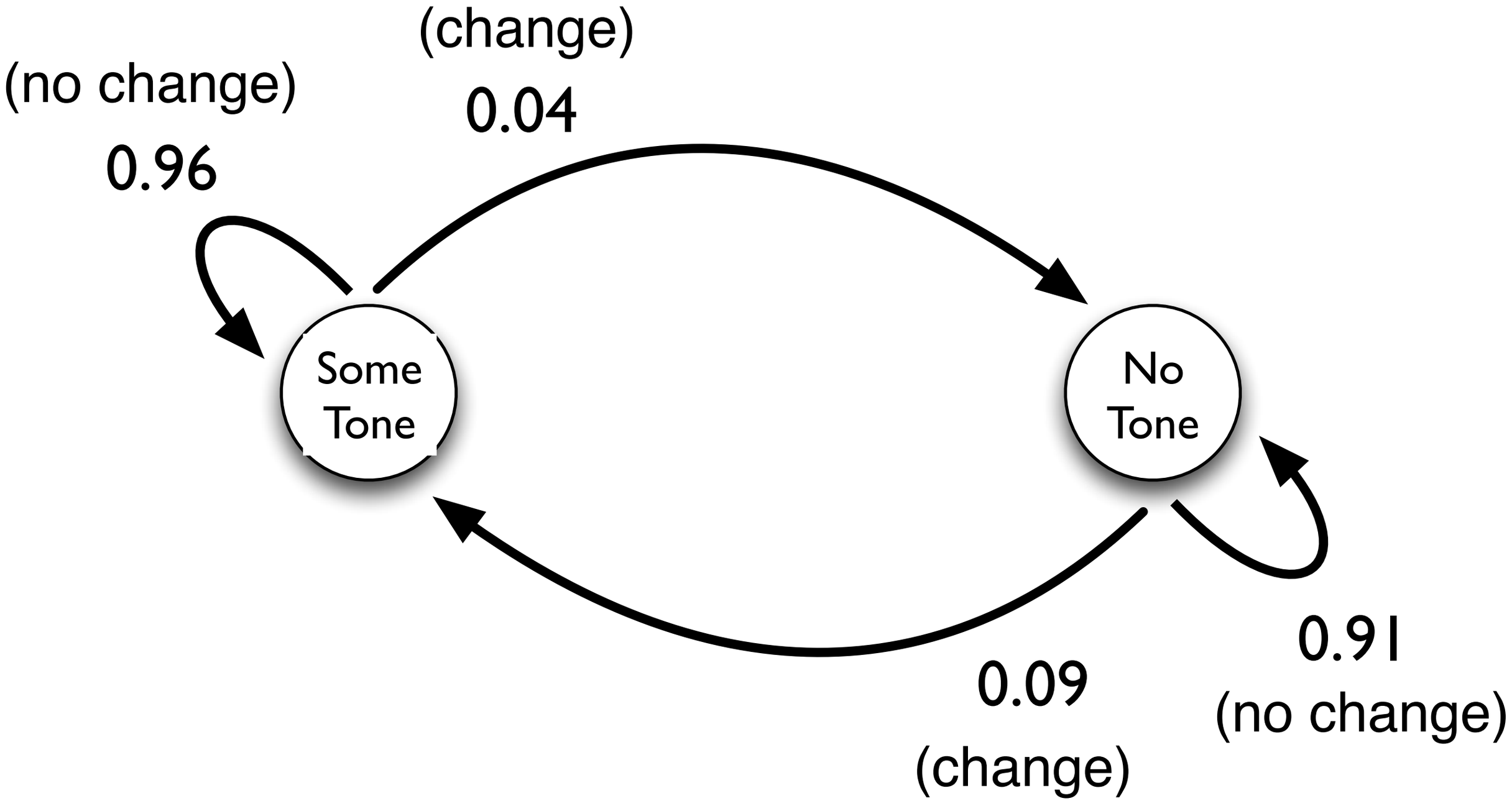

lots of change, so highly variable
Frequency of A << B
little change, so highly stable
1
0
1
$p_{A \to B}$
$p_{B \to A}$

Tone

(Ian Maddieson)




(no change)
1-$p_{A\rightarrow B}$
(change)
$p_{A\rightarrow B}$
Some Tone
No Tone
$p_{B\rightarrow A}$
(change)
1-$p_{B\rightarrow A}$
(no change)

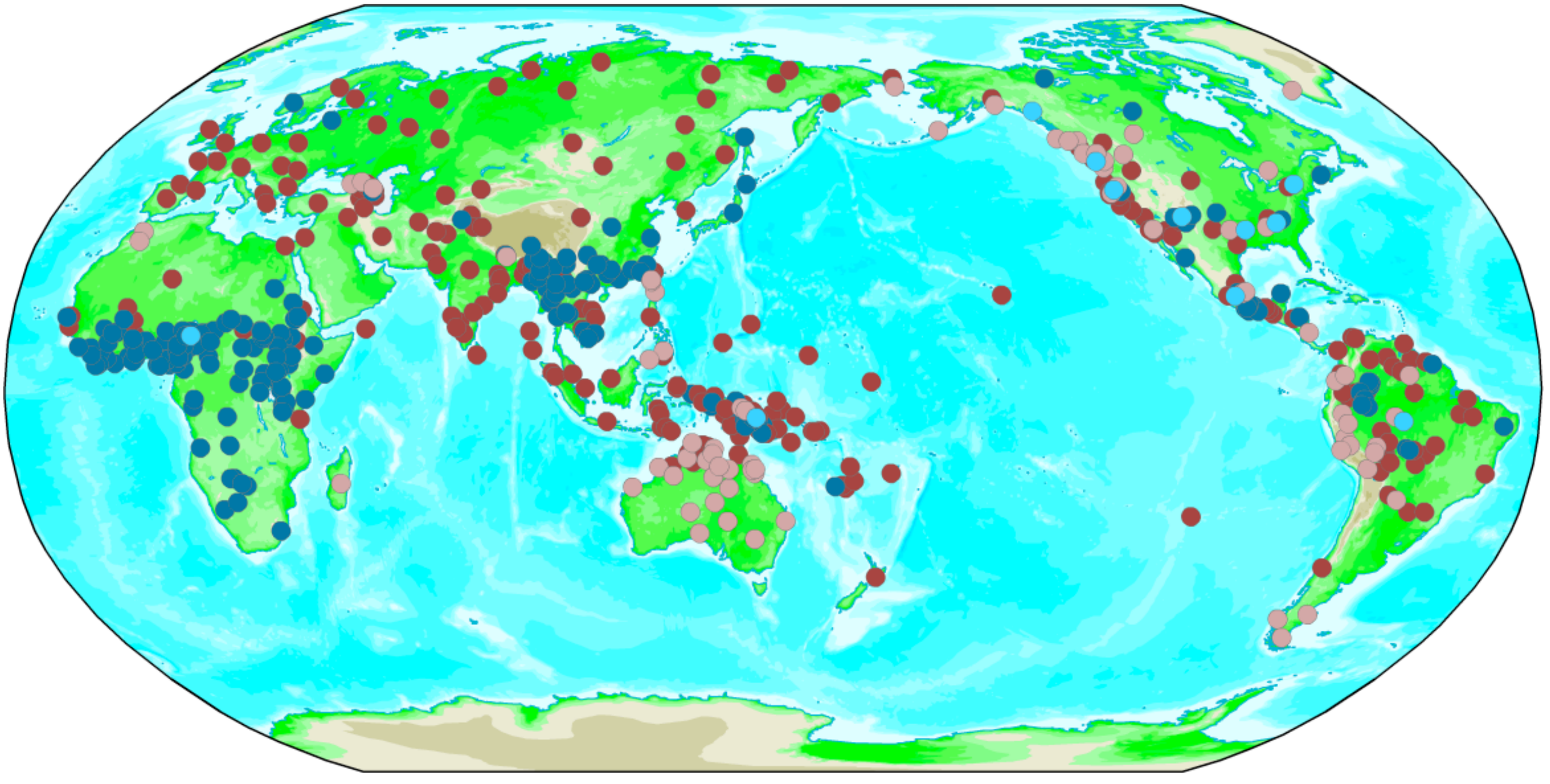

(no change)
0.96
(change)
0.04
Some Tone
No Tone
0.09
(change)
0.91
(no change)

Stable or not ?

	WALS frequency	Expected stable distribution
No tones	306 (58 %)	31 %
Simple tone system	132 (25 %)	23 %
Complex tone system	88 (17 %)	46 %

Cross-section of tone and vowel inventory

(Ian Maddieson)



Traditional Typological Interpretation

	No tone	Tone
Few vowels (<5)	75	11
Many vowels (≥5)	231	206

Tone → Many vowels

Statistical Interpretation

	No tone	Tone
Few vowels (<5)	75	11
Many vowels (≥5)	231	206

$\phi = .26$, Fisher's Exact $p = 7 \cdot 10^{-10}$

Statistical Interpretation

	No tone	Tone
Few vowels (<5)	75 (+25)	11 (-25)
Many vowels (≥5)	231 (-25)	206 (+25)

$\phi = .26$, Fisher's Exact $p = 7 \cdot 10^{-10}$

Tone ~ Many vowels

Dryer's (1992) test

	Africa	Eurasia	SE Asia & Oceania	N. Guinea & Australia	North America	South America
Tone & Large	109	7	41	14	21	14
Tone & Small	1	0	0	1	8	1
No Tone & Large	14	73	44	33	32	35
No Tone & Small	2	3	7	25	21	17

Dryer's (1992) test

	Africa	Eurasia	SE Asia & Oceania	N. Guinea & Australia	North America	South America
Tone & Large	109	7	41	14	21	14
Tone & Small	1	0	0	1	8	1
No Tone & Large	14	73	44	33	32	35
No Tone & Small	2	3	7	25	21	17
p	0,042	n.s.	0,016	0,013	n.s.	0,053

Expected Stable Distribution

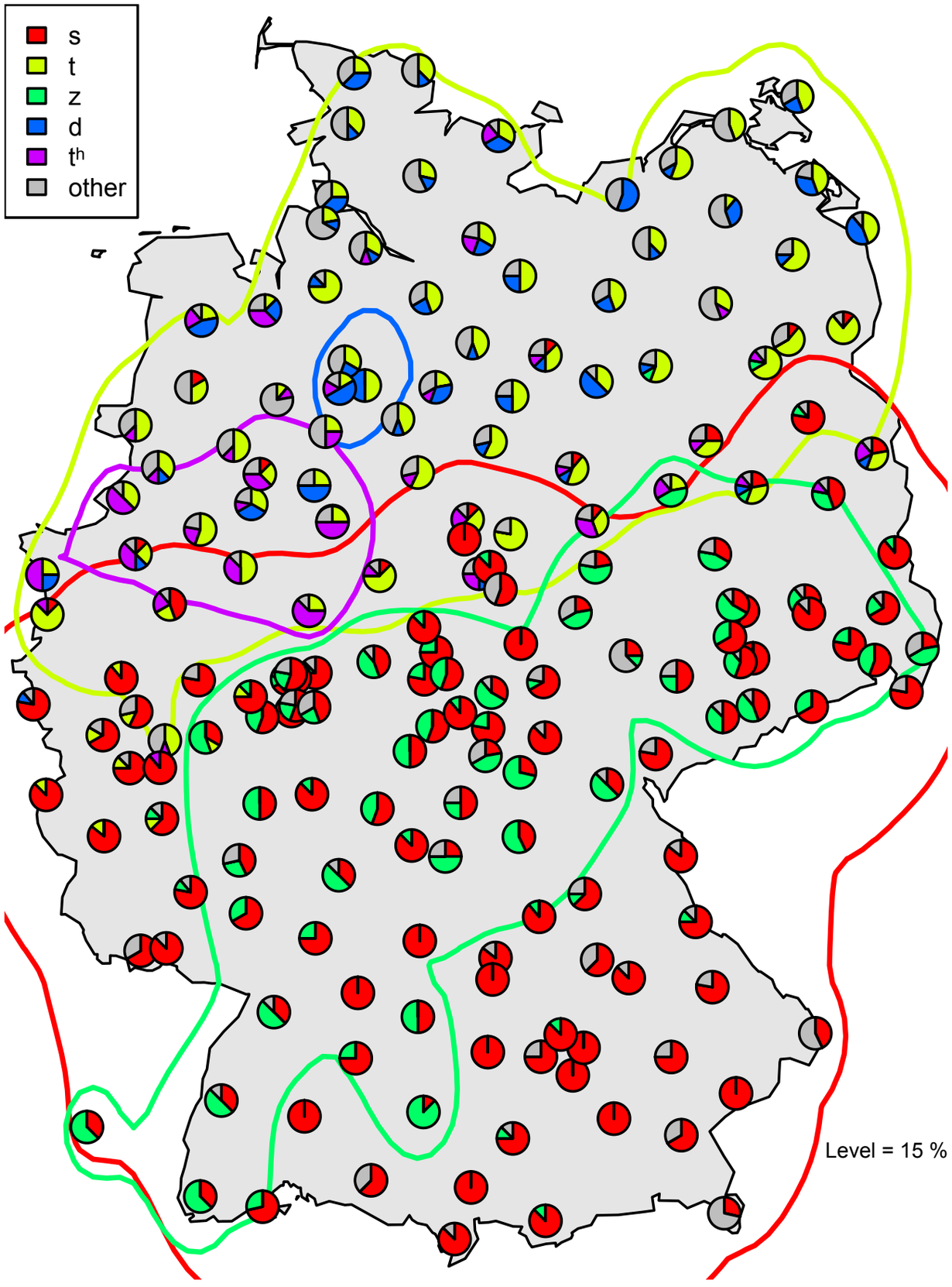
Actual	No tone	Tone
Few vowels (<5)	75	11
Many vowels (≥5)	231	206

Expected Stable Distribution

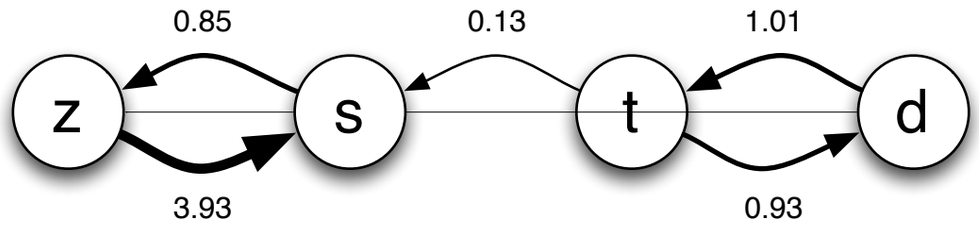
Stable	No tone	Tone
Few vowels (<5)	44	66
Many vowels (≥5)	172	241

ϕ = .01, Fisher's Exact p = .83

Correspondences “s/t”



- Fit a model of sound changes on an unrooted tree based on all correspondences (using *corHMM* in R)
- Continuous-time Markov Chain transition rates:



Based on /s/ in German words:
beißen, besser, das, größer, groß, heiß, muss, Wasser, weiße

Explanation?

- Transition probabilities are no explanation in itself, but 'just' an improved way to report on cross-linguistic distributions
- We need then of course to explain why certain transitions are more probable than others. This is maybe easier than explaining raw frequencies ?
- Tempting idea: maybe these probabilities are also at work at each decision by a speaker to formulate a single utterance