Stable variation arises from noisy across-population bias distributions in the absence of global bias

Henri Kauhanen

The University of Manchester

http://henr.in/

DiGS18 pre-conference workshop
"The Determinants of Diachronic Stability"

Ghent, 28 June 2016

Stable variation arises from noisy across-population bias distributions in the absence of global bias
and from a couple of other things, too

Henri Kauhanen

The University of Manchester

http://henr.in/

DiGS18 pre-conference workshop
"The Determinants of Diachronic Stability"
Ghent, 28 June 2016

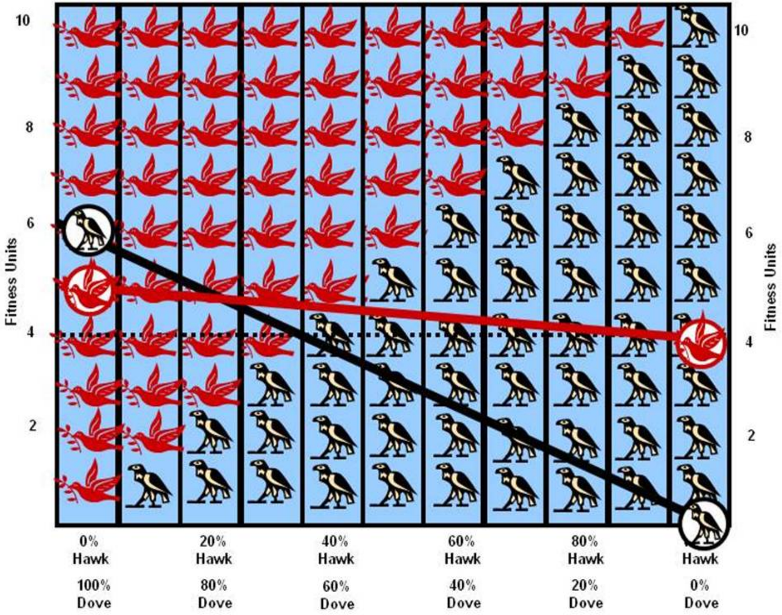
OUTLINE

1. VARIATION AND STABLE VARIATION

- Variation: when more than one variant has non-zero frequency in a population of speakers
- Stable variation: when this state of affairs is strictly stable over time
  - i.e. barring other changes to the system, and discounting stochastic finite-size effects, the state of variation is a stable equilibrium


J.&P. COATS' THREAD
BEST SIX CORD
COTTON
FOR SEWING MACHINES

Fitness achieved in varying populations of Hawk/Dove



2. BIAS AND VARIABLE BIAS

- Now there are mathematical models of change[1,2,3]
- But they mostly assume bias (e.g. phonetic, sociolinguistic) is uniform across speakers
- What if it was variable?

---

[1] Blythe, R. A. & Croft, W. (2012). S-curves and the mechanisms of propagation in language change. *Language*, 88, 269–304.

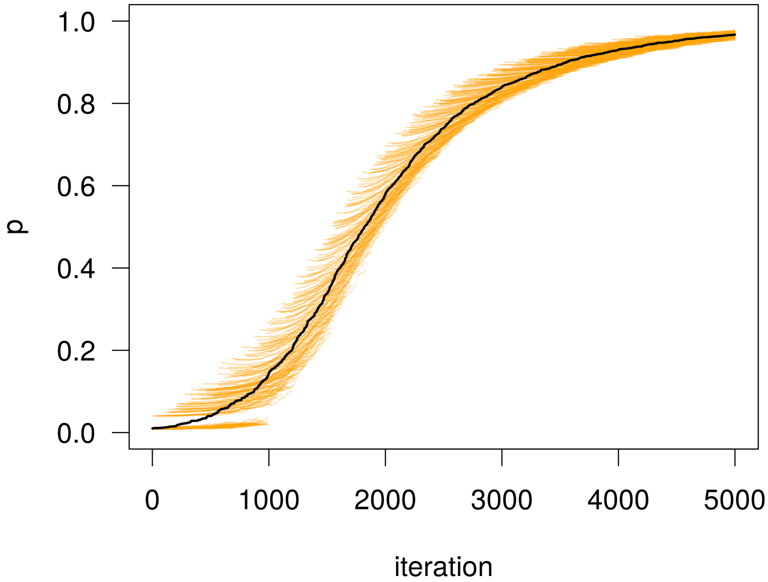
[2] Ke, J., Gong, T. & Wang, W. S.-Y. (2008). Language change and social networks. *Communications in Computational Physics*, 3, 935–949.

[3] Pierrehumbert, J. B. (2001). Exemplar dynamics: word frequency, lenition and contrast. In J. L. Bybee & Paul J. Hopper (Eds.), *Frequency and the emergence of linguistic structure*, 137–157. Amsterdam: Benjamins.

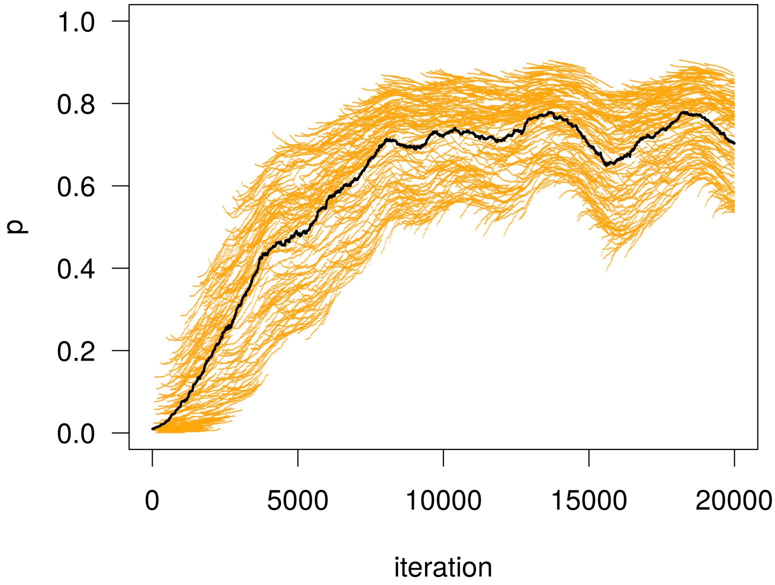
- N speakers
- Binary competition between two variants A and B
- p = prob. of A; $1 - p$ = prob. of B
- b: a bias parameter
- Dynamics: at each iteration, each speaker updates p as

$$p \leftarrow (1-\gamma)p + \gamma \overline{p}^b, \tag{1}$$

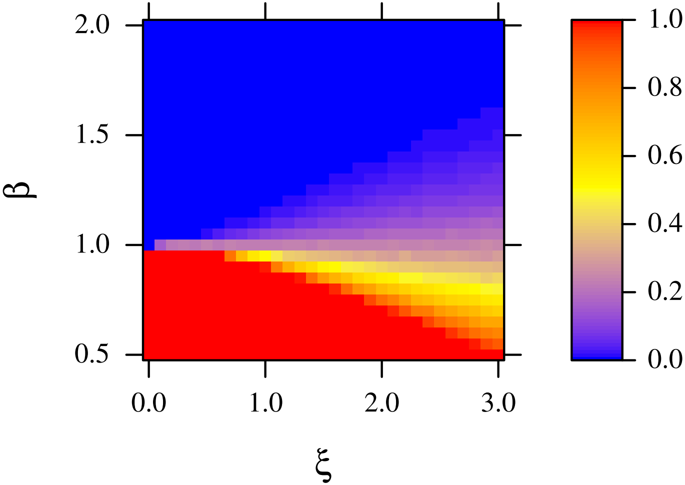
  where
  - $0 < \gamma < 1$ is a learning rate
  - $\overline{p}$ is the mean of p in a random sample of K speakers
- Then $p \to 1$ over time if $0 < b < 1$; and $p \to 0$ if $b > 1$


1.0
0.8
0.6
0.4
0.2
0.0
p
0
1000
2000
3000
4000
5000
iteration

- Now replace the bias parameter b with $\widehat{b}$ sampled uniformly at random from an interval $\left[\frac{\beta}{1+\xi}, \beta + \xi\right]$
- $\widehat{b}$ then varies from speaker to speaker
- Expectation is $E[\widehat{b}] = \beta$
- ξ controls the amount of variation
- Question: how does the asymptotic behaviour of p vary as a function of β and ξ?


1.0
0.8
0.6
0.4
0.2
0.0
p
0
5000
10000
15000
20000
iteration

- Assume definite values for N, K and γ, repeat simulation η times and observe emerging general pattern
  - $\eta = 50$
  - $N = 100$
  - $K = 10$
  - $\gamma = 0.01$
  - assume each speaker lives around 100 iterations
  - β and ξ allowed to vary freely
- Start from an initial state where $\overline{p} = 0.01$



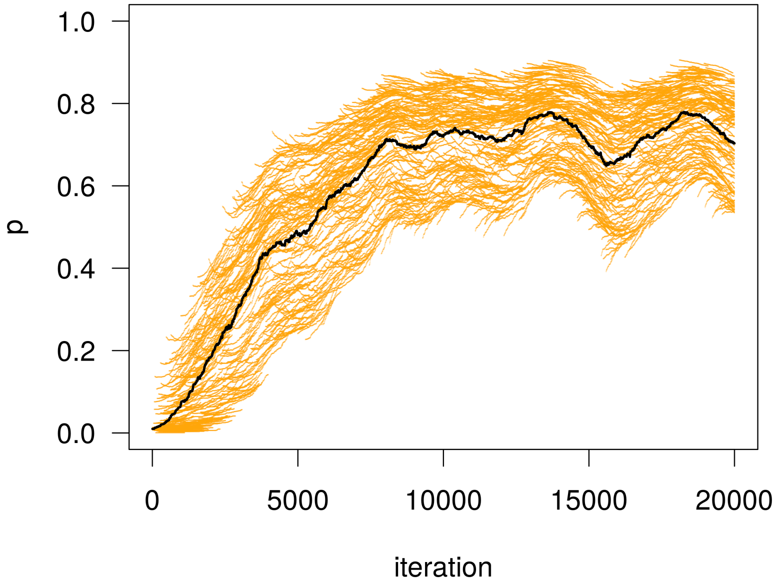
p at 30000 iterations

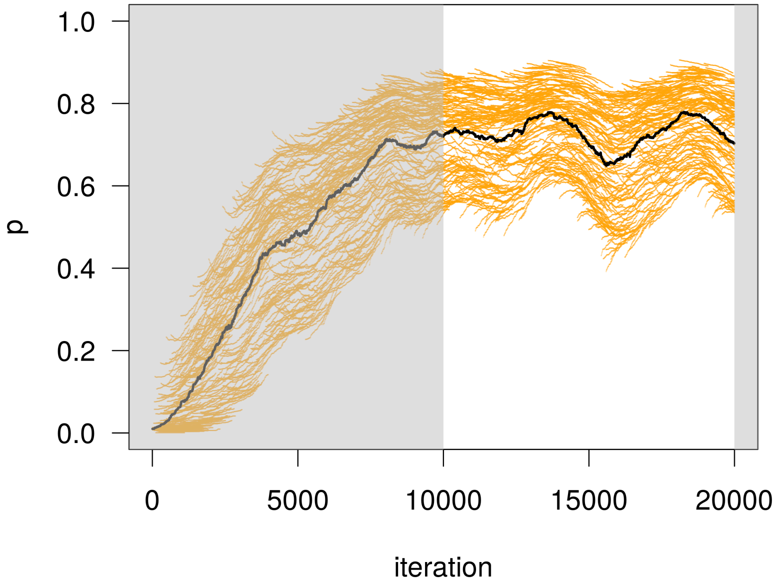
QUANTIFYING STABILITY

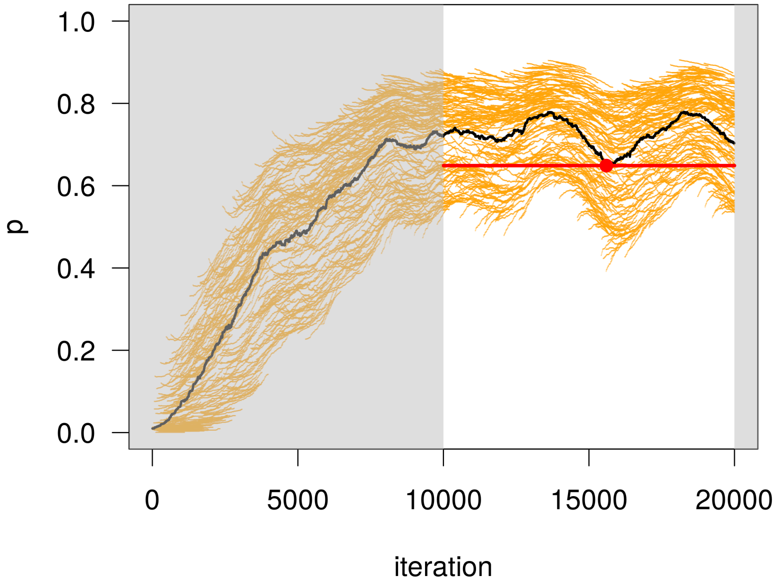
- The above shows that p attains a non-zero, non-unity value if ξ is large enough
- But is this stable?
- Quantify with

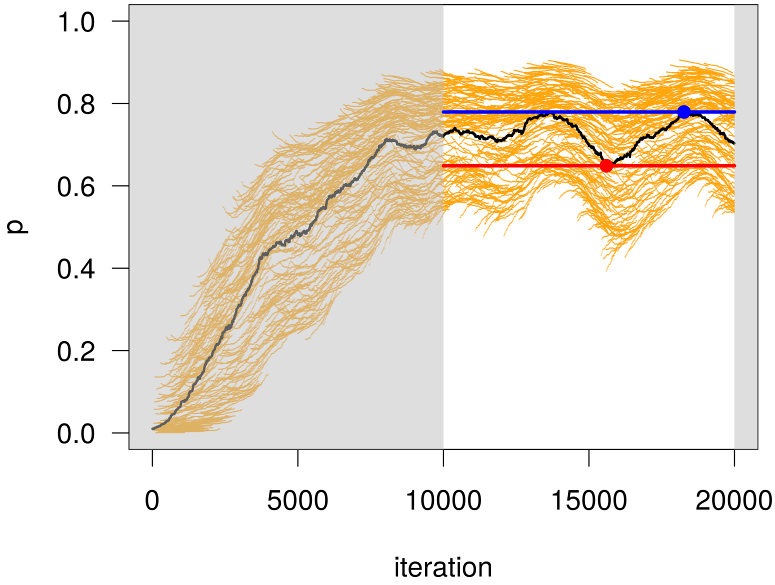
$$S := 1 - (p_{\mathrm{maxfin}} - p_{\mathrm{minfin}}), \tag{2}$$

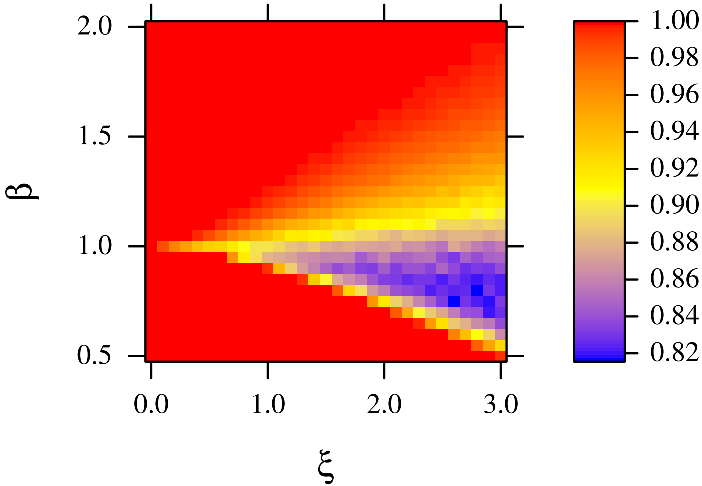
where p_{maxfin} is the maximum of p over the final so-and-so-many iterations and p_{minfin} is the minimum


1.0
0.8
0.6
0.4
0.2
0.0
p
0
5000
10000
15000
20000
iteration


1.0
0.8
0.6
0.4
0.2
0.0
p
0
5000
10000
15000
20000
iteration


1.0
0.8
0.6
0.4
0.2
0.0
p
0
5000
10000
15000
20000
iteration


1.0
0.8
0.6
0.4
0.2
0.0
p
0
5000
10000
15000
20000
iteration



S at 30000 iterations

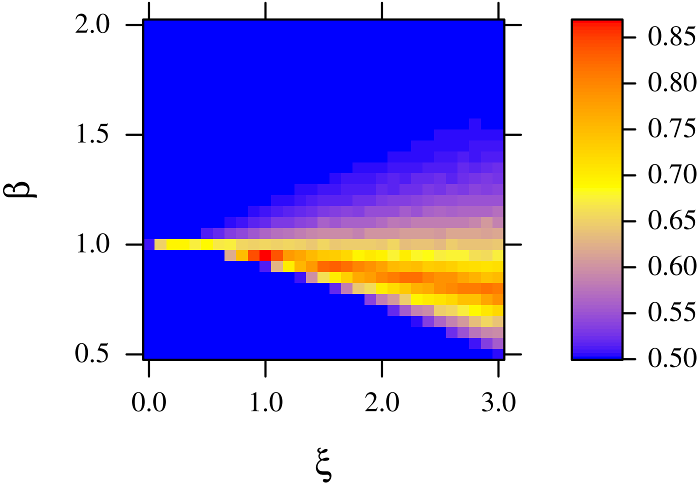
- Finally, let p^* stand for the value of p at the end of the simulation
- We can find out how much p^* diverges from 0.5 (the maximum entropy situation) by considering the inverse distance

$$D := 1 - |0.5 - p^*| \tag{3}$$

- Now consider the product

$$SD = (1 - (p_{\text{maxfin}} - p_{\text{minfin}}))\,(1 - |0.5 - p^*|) \tag{4}$$

- This quantifies, to some extent, the notion of stable *variation*



SD at 30000 iterations

INTERIM SUMMARY

- A mechanism for stable variation: non-uniform bias distributions over speakers in a speech community
- Here, implemented on a *very* simple model
- Need to examine how other types of model will behave under this modification
- Alternative ways of quantifying stability and stable variation could also be explored

3. NON-BINARY COMPETITION

- Another route to stable variation: competition among >2 variants
- Demonstrate this using a generalization of Yang's[4,5] variational learner

[4]Yang, C. D. (2000). Internal and external forces in language change. *Language Variation and Change*, 12, 231–250.

[5]Yang, C. D. (2002). *Knowledge and learning in natural language*. Oxford: Oxford University Press.

- Yang: the learner has access to all grammars licensed by UG and assigns a probability p_i to each grammar G_i with a type of reinforcement learning
- In a monolingual setting learner known to converge to target grammar
- Language change occurs when learner receives input from two grammars G_1 and G_2 and one parses more input than the other

WHAT ABOUT...?

- Question: what might happen in a situation of three-way competition?
  - Input from G_1, G_2 and G_3 available to the learner, with no subset–superset relations among the G_i
- ...in four-way competition?
- ...in n-way competition?

First let's review the 2-grammar case:

1. Learner picks G_1 with prob. p_1 and G_2 with prob. p_2

First let's review the 2-grammar case:

1. Learner picks G_1 with prob. p_1 and G_2 with prob. p_2
2. Assume WLOG G_1 was picked. Present input sentence s.

First let's review the 2-grammar case:

1. Learner picks G_1 with prob. p_1 and G_2 with prob. p_2
2. Assume WLOG G_1 was picked. Present input sentence s.
   - If G_1 parses s, put

$$\begin{cases} p_1 \leftarrow p_1 + \gamma(1 - p_1) \\ p_2 \leftarrow (1 - \gamma)p_2 \end{cases} \tag{5}$$

First let's review the 2-grammar case:

1. Learner picks G_1 with prob. p_1 and G_2 with prob. p_2
2. Assume WLOG G_1 was picked. Present input sentence s.
   - If G_1 parses s, put

$$\begin{cases} p_1 \leftarrow p_1 + \gamma(1 - p_1) \\ p_2 \leftarrow (1 - \gamma)p_2 \end{cases} \tag{5}$$

   - Else, put

$$\begin{cases} p_1 \leftarrow (1 - \gamma)p_1 \\ p_2 \leftarrow p_2 + \gamma(1 - p_2) \end{cases} \tag{6}$$

where $0 < \gamma < 1$ is a learning rate.

First let's review the 2-grammar case:

1. Learner picks G_1 with prob. p_1 and G_2 with prob. p_2
2. Assume WLOG G_1 was picked. Present input sentence s.
   - If G_1 parses s, put

$$\begin{cases} p_1 \leftarrow p_1 + \gamma(1 - p_1) \\ p_2 \leftarrow (1 - \gamma)p_2 \end{cases} \tag{5}$$

   - Else, put

$$\begin{cases} p_1 \leftarrow (1 - \gamma)p_1 \\ p_2 \leftarrow p_2 + \gamma(1 - p_2) \end{cases} \tag{6}$$

   where $0 < \gamma < 1$ is a learning rate.
3. This is iterated for N input sentences.

First let's review the 2-grammar case:

1. Learner picks G_1 with prob. p_1 and G_2 with prob. p_2
2. Assume WLOG G_1 was picked. Present input sentence s.
   - If G_1 parses s, put

$$\begin{cases} p_1 \leftarrow p_1 + \gamma(1 - p_1) \\ p_2 \leftarrow (1 - \gamma)p_2 \end{cases} \tag{7}$$

   - Else, put

$$\begin{cases} p_1 \leftarrow (1 - \gamma)p_1 \\ p_2 \leftarrow p_2 + \gamma(1 - p_2) \end{cases} \tag{8}$$

   where $0 < \gamma < 1$ is a learning rate.
3. This is iterated for N input sentences.

First let's review the 2-grammar case:

1. Learner picks G_1 with prob. p_1 and G_2 with prob. p_2
2. Assume WLOG G_1 was picked. Present input sentence s.
   - If G_1 parses s, put

$$\begin{cases} p_1 \leftarrow p_1 + \gamma(1 - p_1) \\ p_2 \leftarrow (1 - \gamma)p_2 \end{cases} \tag{9}$$

   - Else, put

$$\begin{cases} p_1 \leftarrow (1 - \gamma)p_1 \\ p_2 \leftarrow p_2 + \gamma(1 - p_2) \end{cases} \tag{10}$$

   where $0 < \gamma < 1$ is a learning rate.
3. This is iterated for N input sentences.

- The **penalty probability** of grammar G_i, relative to environment E, is

$$c_i = \text{Prob}(G_i \text{ doesn't parse } s \mid s \in E) \tag{11}$$

- The **penalty probability** of grammar G_i, relative to environment E, is

$$c_i = \text{Prob}(G_i \text{ doesn't parse } s \mid s \in E) \tag{11}$$

- Now notice: in a 2-grammar setting, the penalty of G_1 can be expressed as

$$c_1 = \alpha_2 \pi_2, \tag{12}$$

  where
  - π_2 = prob. that the learner encounters a sentence generated by G_2
  - α_2 = prob. that this sentence is *not* parsable by G_1 (called the **advantage** of G_2 by Yang)
- By symmetry, $c_2 = \alpha_1 \pi_1$

Theorem (Narendra & Thathachar[6])

Assume the LRP learner samples a stationary environment at random. Then, if $N\gamma \to \infty$ and $\gamma \to 0$, the weights are normally distributed with means satisfying

$$E[p_1] \to \frac{1}{1 + c_1/c_2} \quad \text{and} \quad E[p_2] \to \frac{1}{1 + c_2/c_1} \tag{13}$$

and variances which tend to 0.

- In other words: supposing the learner receives a large number of input sentences (large N) and only makes small adjustments (small γ), he will settle upon weights which are very well approximated by

$$p_1^* = \frac{1}{1 + \frac{\alpha_2 \pi_2}{\alpha_1 \pi_1}} \quad \text{and} \quad p_2^* = \frac{1}{1 + \frac{\alpha_1 \pi_1}{\alpha_2 \pi_2}} \tag{14}$$

[6] Narendra, K. & Thathachar, M. A. L. (1989). *Learning automata: an introduction*. Englewood Cliffs, NJ: Prentice-Hall.

- Assuming non-overlapping generations of learners, this gives the inter-generational dynamics

$$\begin{cases} p_1(t+1) = \left(1 + \frac{\alpha_2 p_2(t)}{\alpha_1 p_1(t)}\right)^{-1} \\ p_2(t+1) = \left(1 + \frac{\alpha_1 p_1(t)}{\alpha_2 p_2(t)}\right)^{-1} \end{cases} \tag{15}$$

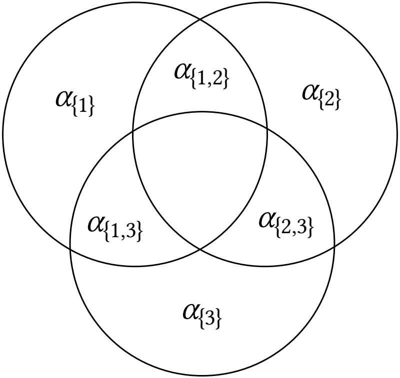
- Logistic with slope governed by the ratio α_1/α_2:
  - G_1 wins if $\alpha_1 > \alpha_2$
  - G_2 wins if $\alpha_1 < \alpha_2$

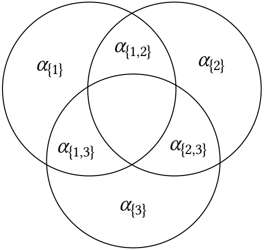
- Assuming non-overlapping generations of learners, this gives the inter-generational dynamics

$$\begin{cases} p_1(t+1) = \left(1 + \frac{\alpha_2 p_2(t)}{\alpha_1 p_1(t)}\right)^{-1} \\ p_2(t+1) = \left(1 + \frac{\alpha_1 p_1(t)}{\alpha_2 p_2(t)}\right)^{-1} \end{cases} \tag{15}$$

- Logistic with slope governed by the ratio α_1/α_2:
  - G_1 wins if $\alpha_1 > \alpha_2$
  - G_2 wins if $\alpha_1 < \alpha_2$
- Therefore predicts stable variation to be **impossible** (this is Yang's Fundamental Theorem of Language Change)

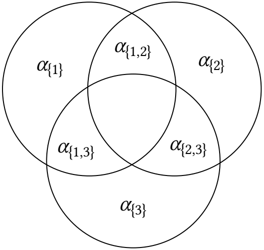
- The LRP algorithm works for an arbitrary number n of grammars
- A derivation analogous to the above, but more complicated, obtains. Consider $n = 3$ first:





- For G_1, one has the penalty

$$c_1 = \alpha_{\{2\}} p_2$$

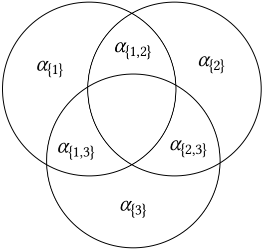


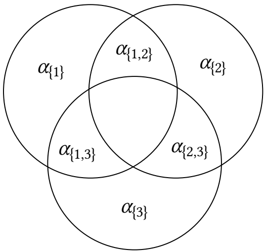
- For G_1, one has the penalty

$$c_1 = \alpha_{\{2\}} p_2 + \alpha_{\{3\}} p_3$$

- For G_1, one has the penalty

$$c_1 = \alpha_{\{2\}} p_2 + \alpha_{\{3\}} p_3 + \alpha_{\{2,3\}} (p_2 + p_3) \quad (16)$$





- For G_1, one has the penalty

$$c_1 = \alpha_{\{2\}} p_2 + \alpha_{\{3\}} p_3 + \alpha_{\{2,3\}}(p_2 + p_3) \quad (16)$$

Rearrange terms:

$$(\underbrace{\alpha_{\{2\}} + \alpha_{\{2,3\}}}_{=:\alpha_{21}})p_2 + (\underbrace{\alpha_{\{3\}} + \alpha_{\{2,3\}}}_{=:\alpha_{31}})p_3 \quad (17)$$

and call $\alpha_{21} = \alpha_{\{2\}} + \alpha_{\{2,3\}}$ the **relative advantage** of G_2 over G_1 (similarly for α_{31})

- Intuitively, α_{ji} = prob. of a sentence which is parsed by G_j but not by G_i

- Procedure generalizes to n grammars, whereby one finds

$$c_i = \sum_{\substack{j=1 \\ j \neq i}}^{n} \alpha_{ji} p_j \tag{18}$$

with

$$\alpha_{ji} = \sum_{K \in \mathcal{K}_{ji}} \alpha_K, \tag{19}$$

where $\mathcal{K}_{ji} = \{X \subseteq \{1, \ldots, n\} \setminus \{i\} \mid j \in X\}$.

- Here α_K = prob. of a sentence which is parsed by all and only the grammars G_ℓ with $\ell \in K$

- Now collect the relative advantages in an **advantage matrix**:

$$A = \begin{pmatrix} 0 & \alpha_{12} & \dots & \alpha_{1n} \\ \alpha_{21} & 0 & \dots & \alpha_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{n1} & \alpha_{n2} & \dots & 0 \end{pmatrix} \tag{20}$$

- The properties of this matrix will determine the dynamics
- System still very simple, so can be explored analytically
- Strategy: start with simple advantage matrices, proceeding then to more complicated cases

- Inter-generational evolution equation now becomes

$$p_i(t+1) = \left(1 + \sum_{\substack{j=1 \\ j\neq i}}^{n} \frac{\sum_{k=1}^{n} \alpha_{ki} p_k(t)}{\sum_{\ell=1}^{n} \alpha_{\ell j} p_\ell(t)}\right)^{-1} \tag{21}$$

  for $i = 1, \ldots, n$.

- For $n = 3$, we have

$$\begin{cases} p_1(t+1) = \left(1 + \frac{\alpha_{21}p_2(t)+\alpha_{31}p_3(t)}{\alpha_{12}p_1(t)+\alpha_{32}p_3(t)} + \frac{\alpha_{21}p_2(t)+\alpha_{31}p_3(t)}{\alpha_{13}p_1(t)+\alpha_{23}p_2(t)}\right)^{-1} \\ p_2(t+1) = \left(1 + \frac{\alpha_{12}p_1(t)+\alpha_{32}p_3(t)}{\alpha_{21}p_2(t)+\alpha_{31}p_3(t)} + \frac{\alpha_{12}p_1(t)+\alpha_{32}p_3(t)}{\alpha_{23}p_2(t)+\alpha_{13}p_1(t)}\right)^{-1} \\ p_3(t+1) = \left(1 + \frac{\alpha_{13}p_1(t)+\alpha_{23}p_2(t)}{\alpha_{31}p_3(t)+\alpha_{21}p_2(t)} + \frac{\alpha_{13}p_1(t)+\alpha_{23}p_2(t)}{\alpha_{32}p_3(t)+\alpha_{12}p_1(t)}\right)^{-1} \end{cases} \tag{22}$$

- Let x denote the state of the system at an arbitrary time, and let $f^t(x)$ be the state after t iterations

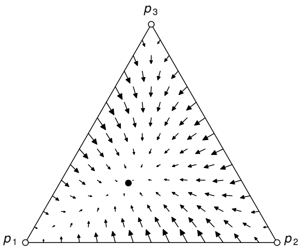
- Let x denote the state of the system at an arbitrary time, and let $f^t(x)$ be the state after t iterations
- Fix(ed)point or equilibrium: a state x with the property $f^1(x) = x$

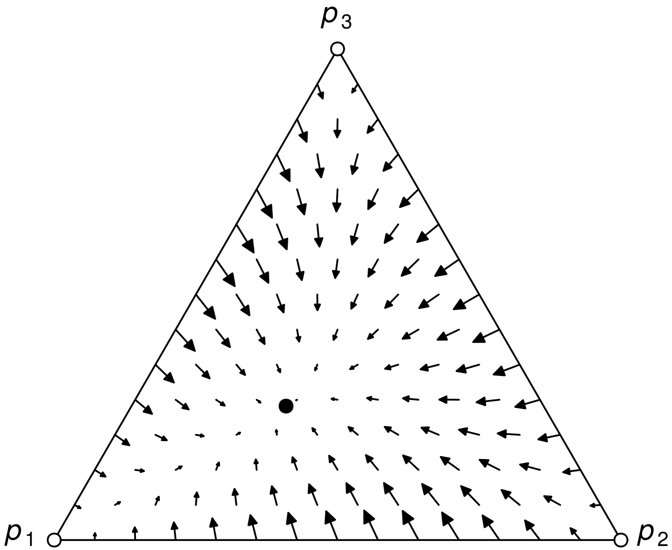
- Let x denote the state of the system at an arbitrary time, and let $f^t(x)$ be the state after t iterations
- Fix(ed)point or equilibrium: a state x with the property $f^1(x) = x$
- Locally stable fixpoint: a fixpoint x with this property: for any state y sufficiently close to x, $f^t(y) \to x$ as $t \to \infty$

- Let x denote the state of the system at an arbitrary time, and let $f^t(x)$ be the state after t iterations
- Fix(ed)point or equilibrium: a state x with the property $f^1(x) = x$
- Locally stable fixpoint: a fixpoint x with this property: for any state y sufficiently close to x, $f^t(y) \to x$ as $t \to \infty$
- Globally stable fixpoint: a fixpoint x with this property: for *any* state y, $f^t(y) \to x$ as $t \to \infty$

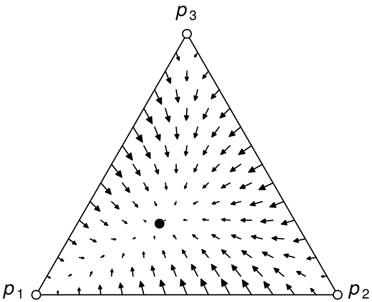
- Let x denote the state of the system at an arbitrary time, and let $f^t(x)$ be the state after t iterations
- Fix(ed)point or equilibrium: a state x with the property $f^1(x) = x$
- Locally stable fixpoint: a fixpoint x with this property: for any state y sufficiently close to x, $f^t(y) \to x$ as $t \to \infty$
- Globally stable fixpoint: a fixpoint x with this property: for *any* state y, $f^t(y) \to x$ as $t \to \infty$
- Unstable fixpoint: a fixpoint x that is neither locally nor globally stable

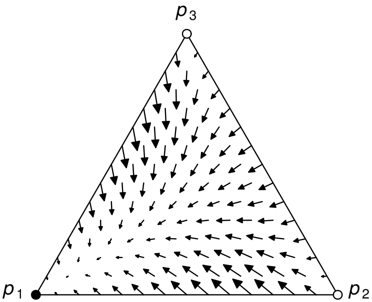
- For us, the system states are vectors of grammar weights: $x = \mathbf{p} = (p_1, p_2, \ldots, p_n)$
- Stable variation, then, occurs if we have a stable fixpoint $\mathbf{p}$ with this property: $p_i = 1$ for *no* grammar G_i
- For $n = 3$, best illustrated using a **ternary plot**:




p_3
p_1
p_2

- Easily proved: the **vertices** (states where $p_i = 1$ for some i) are fixpoints for *any* advantage matrix A
- Their stability, however, may change as A changes

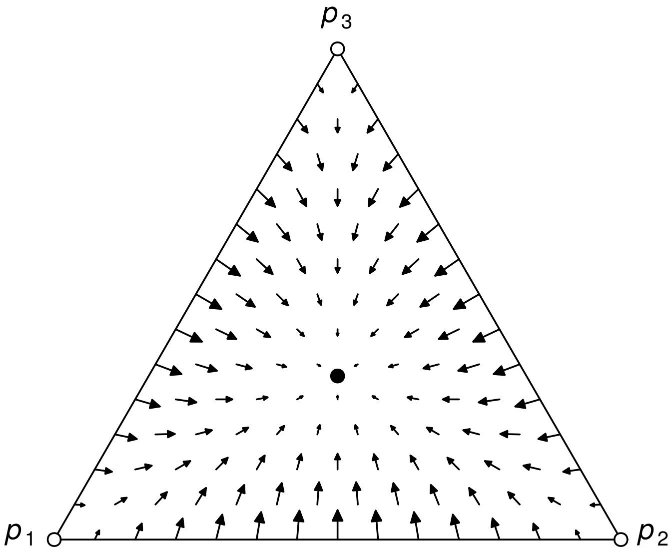




- Start with the simplest system possible: no asymmetries in pairwise grammar advantages
- I.e. advantage matrix is of the form

$$A = \begin{pmatrix} 0 & a & \dots & a \\ a & 0 & \dots & a \\ \vdots & \vdots & \ddots & \vdots \\ a & a & \dots & 0 \end{pmatrix} \tag{23}$$

  for some a. Call such a system **Babelian**.
- Result:
  1. The vertex fixpoints are **unstable**
  2. There is a single **globally stable** fixpoint at $\mathbf{p} = (1/n, 1/n, \dots, 1/n)$


p_3
p_1
p_2

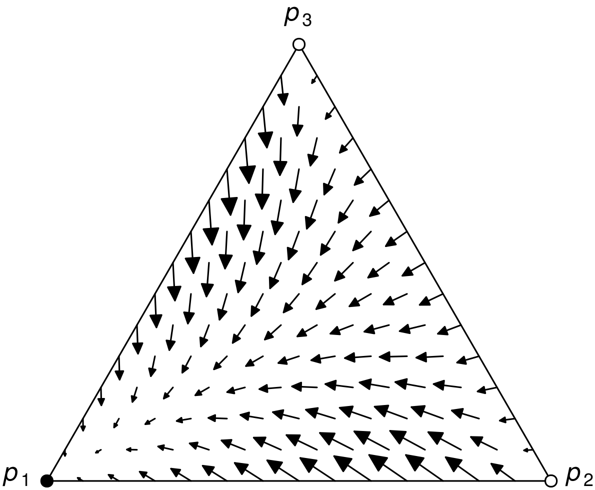
- Now assume advantage matrix is of the form

$$A = \begin{pmatrix} 0 & b & \dots & b \\ a & 0 & \dots & a \\ \vdots & \vdots & \ddots & \vdots \\ a & a & \dots & 0 \end{pmatrix} \tag{24}$$

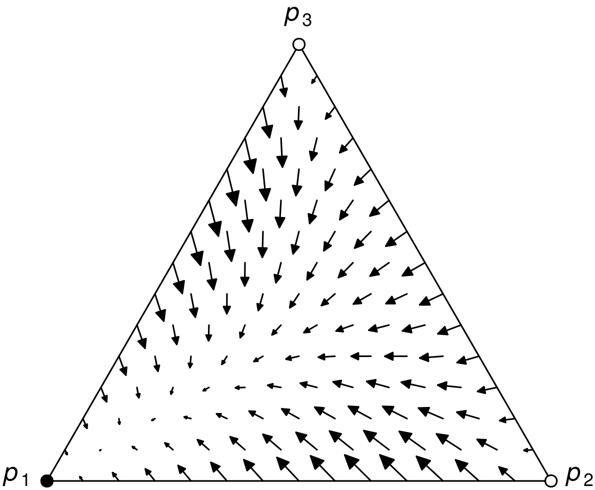
  i.e. Babelian up to G_1 (WLOG), which has b rather than a
- Call such a system **semi-Babelian**
- If $b > a$, G_1 is "better" than the rest
- If $b < a$, the rest are "better" than G_1

- Behaviour of such a system is more complicated, with b/a a bifurcation parameter
- Can be exactly solved in the 3-grammar case, where we have:
  - Assume $b/a \geq 2$. Then only the vertex fixpoints exist and they are **stable**
  - Assume $0 \leq b/a < 2$. Then the vertex fixpoints are **unstable** and a further **stable** fixpoint exists at

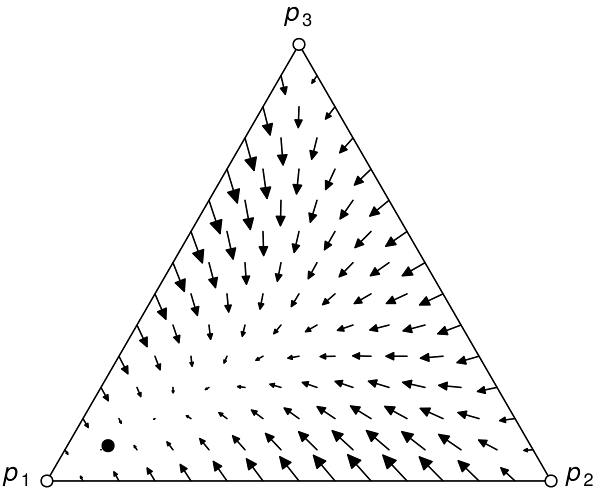
$$(p_1, p_2, p_3) = \left(-\frac{1}{2b/a - 5}, \frac{b/a - 2}{2b/a - 5}, \frac{b/a - 2}{2b/a - 5}\right) \tag{25}$$



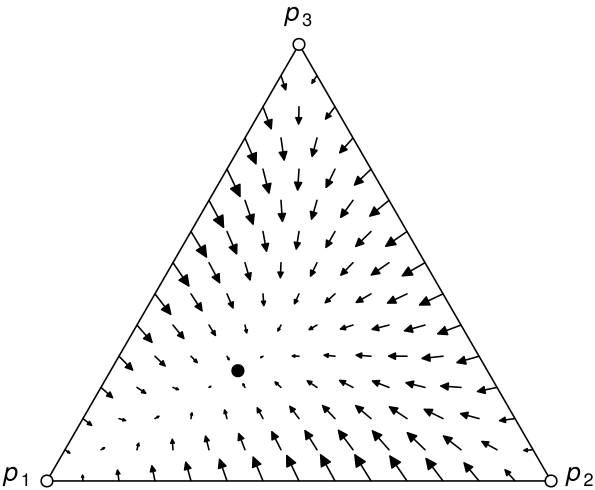
$b/a = 3.0$



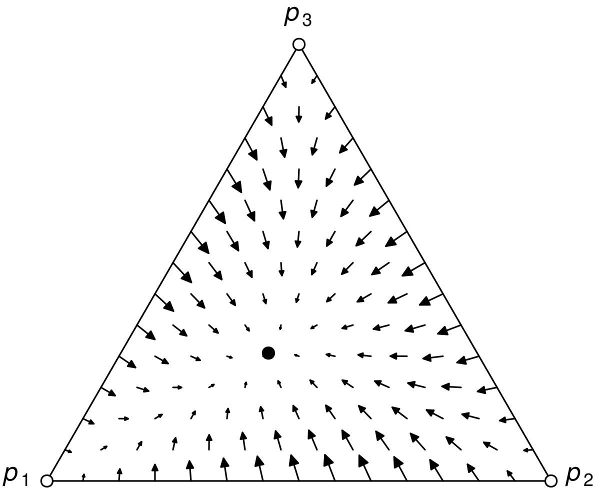
$b/a = 2.1$



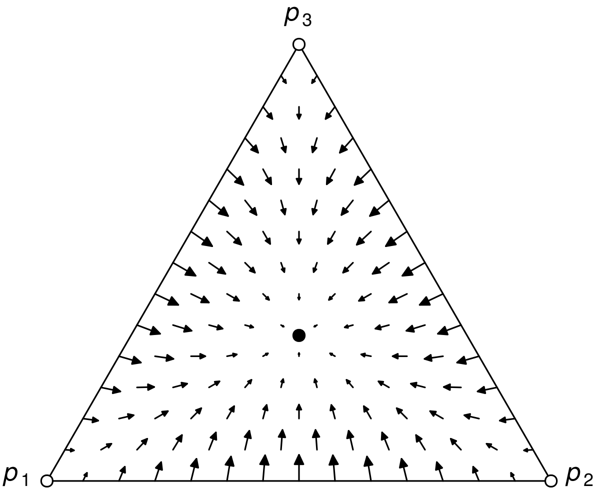
$b/a = 1.9$



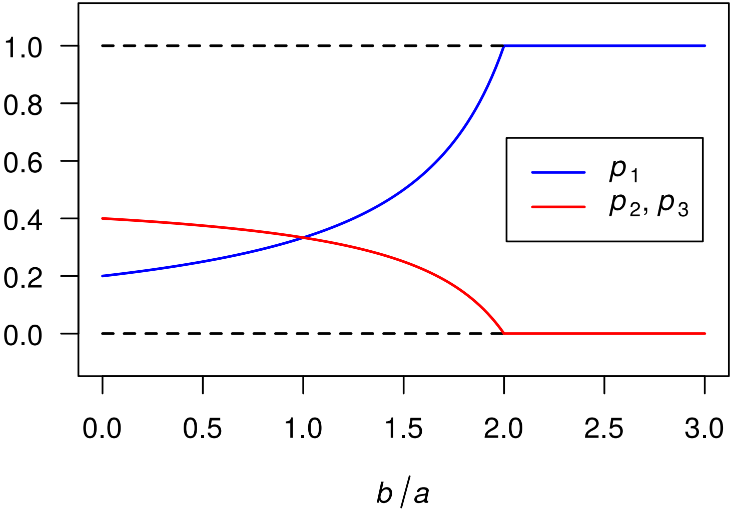
$b/a = 1.5$



$b/a = 1.3$



$b/a = 1.0$


1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
b / a
p 1
p 2, p 3

- Have generalized Yang's model of language change to a setting of n-way competition
- Have abstracted from this the two very special cases of situation: Babelian and semi-Babelian
- Have demonstrated that stable variation occurs in both cases
  - In Babelian systems: always
  - In semi-Babelian systems: if one grammar is not too advantageous compared to the rest
- Empirical work needs to establish what this means
  - Does n-way competition with $n > 2$ actually exist?
  - If so, do we see a tendency towards stable variation?

4. PARAMETRIC SPACES

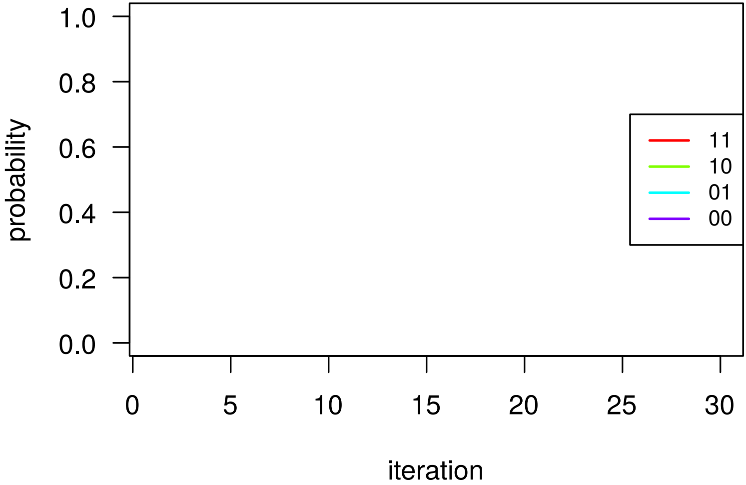
- Problem: learners operate in a parametric space
- I.e. it does not seem likely that the learner assigns a probability to each grammar (of which there are astronomically many!), but to a number of syntactic parameters
- Does this affect our results?

NAIVE PARAMETER LEARNER (NPL)

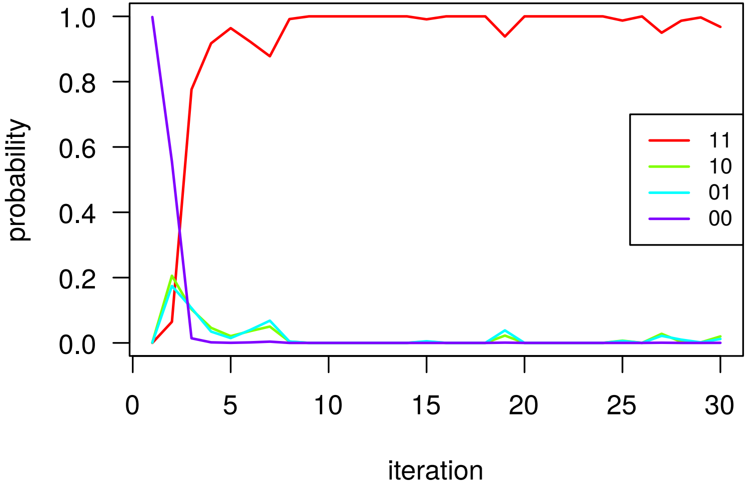
- So now, assume learner sets probabilities p_i for **syntactic parameters**, not for grammars
  - Assume binary parameters: then n parameters give 2^n grammars
- Problem: if parsing event is unsuccessful, how does learner know which parameter setting was the culprit?
- Yang's[7] **Naive Parameter Learner**: demote *all* parameter probabilities with unsuccessful parsing, promote *all* parameter probabilities with successful parsing
- Problem: Narendra & Thathachar's asymptotic result on LRP learning not available
  - Response: fall back on computer simulations, assuming parameter independence for simplicity

[7] Yang, C. D. (2002). *Knowledge and learning in natural language*. Oxford: Oxford University Press.

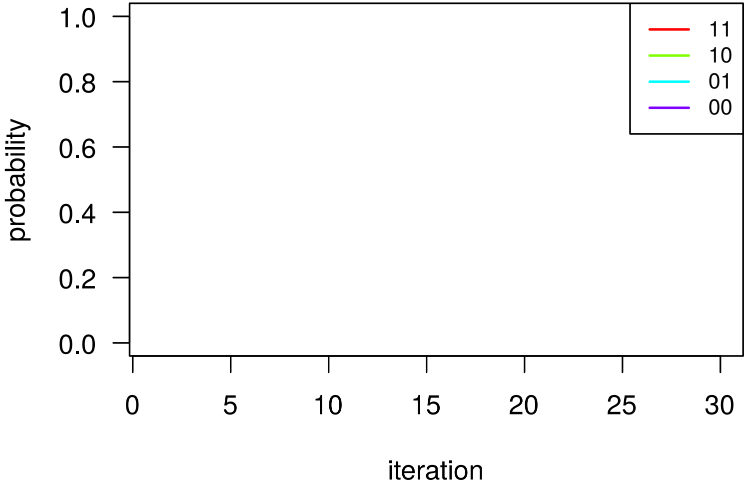
- Computer simulations of iterated NPL with following model parameters:
  - $n = 2$: number of syntactic parameters (thus 4 grammars)
  - $N = 20000$: number of sentences learner hears
  - $\gamma = 0.005$: learning rate, as in LRP
  - α_i: prob. of a sentence which is only parsable if the ith parameter is set **on**
  - β_i: prob. of a sentence which is only parsable if the ith parameter is set **off**



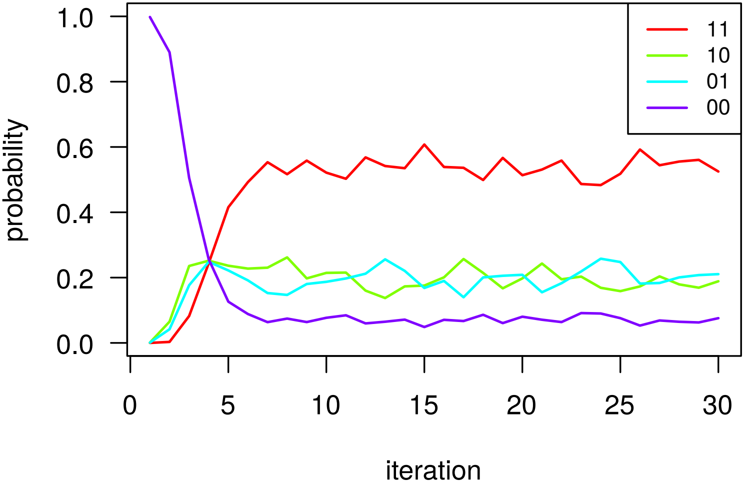
$\alpha_1 = \alpha_2 = 0.9, \quad \beta_1 = \beta_2 = 0.01$



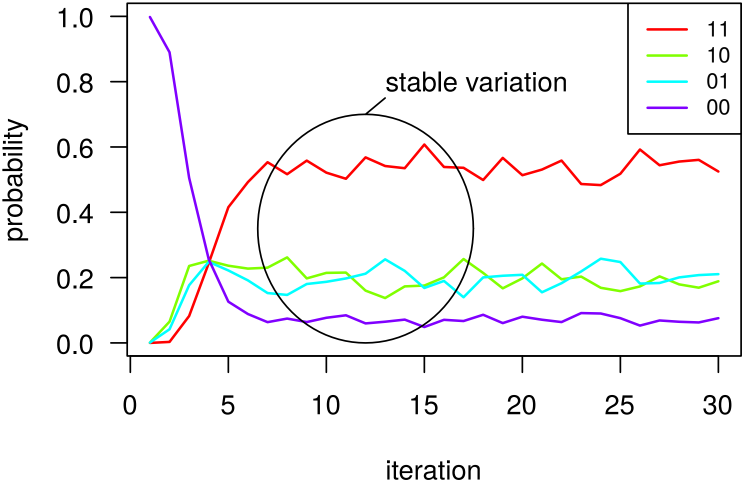
$\alpha_1 = \alpha_2 = 0.9, \quad \beta_1 = \beta_2 = 0.01$



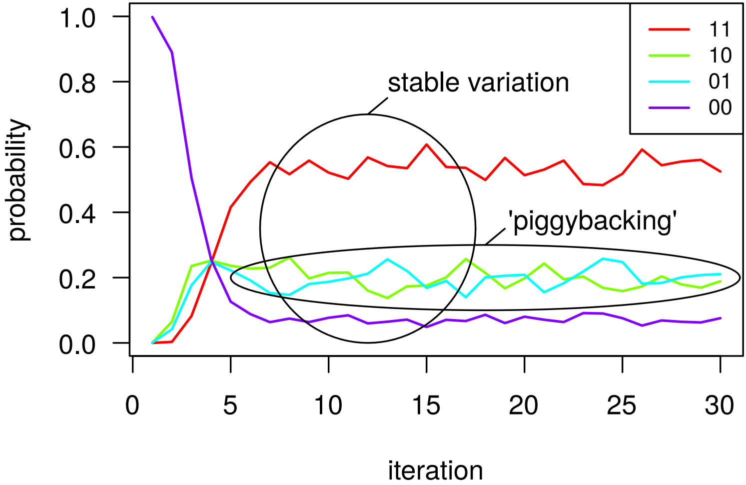
$\alpha_1 = \alpha_2 = 0.3, \quad \beta_1 = \beta_2 = 0.1$



$\alpha_1 = \alpha_2 = 0.3, \quad \beta_1 = \beta_2 = 0.1$



$\alpha_1 = \alpha_2 = 0.3, \quad \beta_1 = \beta_2 = 0.1$



$\alpha_1 = \alpha_2 = 0.3, \quad \beta_1 = \beta_2 = 0.1$

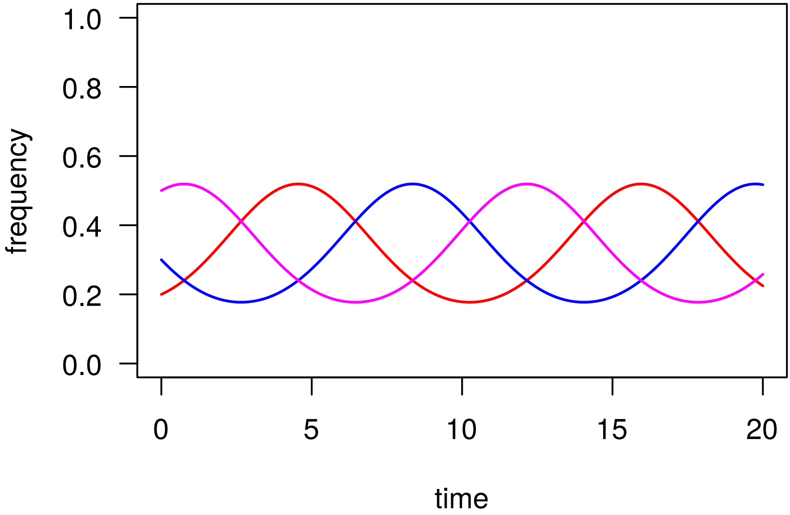
5. CONCLUSIONS

- We have defined stable variation as a state **p** with the following properties:
  1. no variant has probability one, i.e. $p_i = 1$ for *no i*
  2. it is asymptotically stable: barring a tweaking of the system's parameters, and discounting stochastic finite-size effects, the system's state will be attracted to **p** over positive time

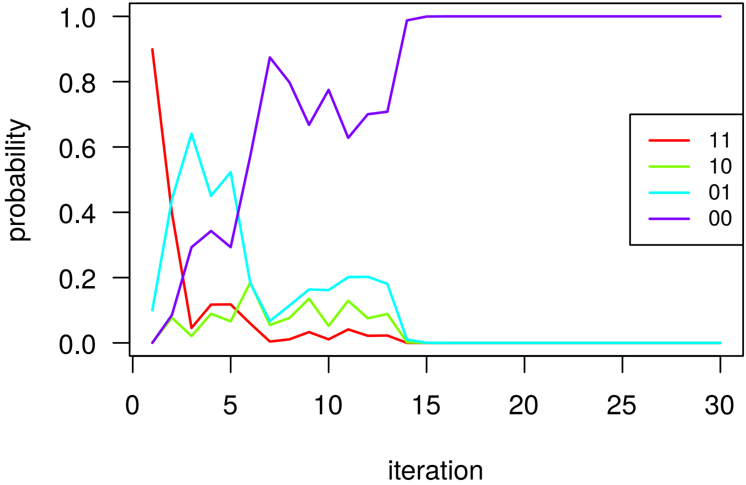
  A **point attractor** in Dyn Syst terminology.
- We have shown that such states can exist in at least three ways:
  - Inter-speaker variation in the way a bias is applied
  - Intra-speaker distributions of > 2 grammar probabilities
  - Intra-speaker distributions of > 1 parameter probabilities

SUMMARY

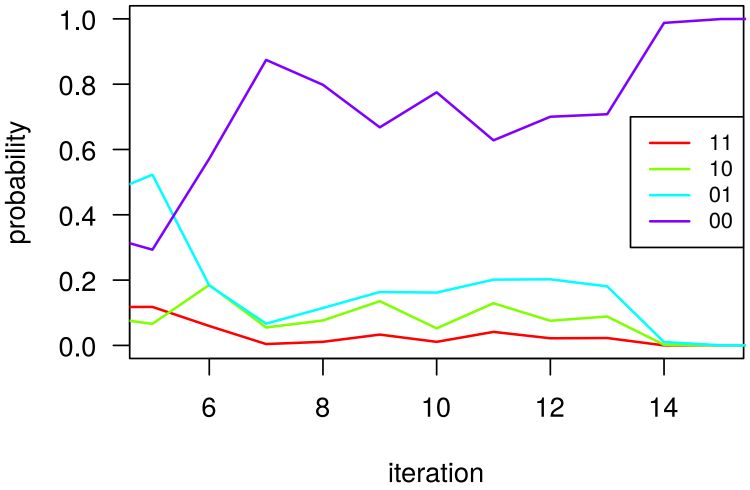
- The argument here is a conceptual one: this is what these models predict
- Stable variation is **endemic** to the models: if the models capture reality at all, we should see plenty of stable variation in the real world
- Need to explore this in more detail from an empirical point of view
- Conceptual issues that need some thought:
  - Does it have to be a *point* attractor?


1.0
0.8
0.6
0.4
0.2
0.0
frequency
0
5
10
15
20
time

- The argument here is a conceptual one: this is what these models predict
- Stable variation is **endemic** to the models: if the models capture reality at all, we should see plenty of stable variation in the real world
- Need to explore this in more detail from an empirical point of view
- Conceptual issues that need some thought:
  - Does it have to be a *point* attractor?
  - How to think about finite-size effects?



$\alpha_1 = \alpha_2 = 0.1, \quad \beta_1 = \beta_2 = 0.4, \quad N = 5000, \quad \gamma = 0.05$



$\alpha_1 = \alpha_2 = 0.1, \quad \beta_1 = \beta_2 = 0.4, \quad N = 5000, \quad \gamma = 0.05$

- George Walkden
- Ricardo Bermúdez-Otero
- Emil Aaltonen Foundation