LI 511: Computational Models of Sound Change

James Kirby and Morgan Sonderegger

5 July 2013

Administrative

- Office hours: Monday
  - 9-11? 1:30-3:30?
  - (or by appointment!)

- Final project
  - Fri next week: short email (250 word max) on topic + plan

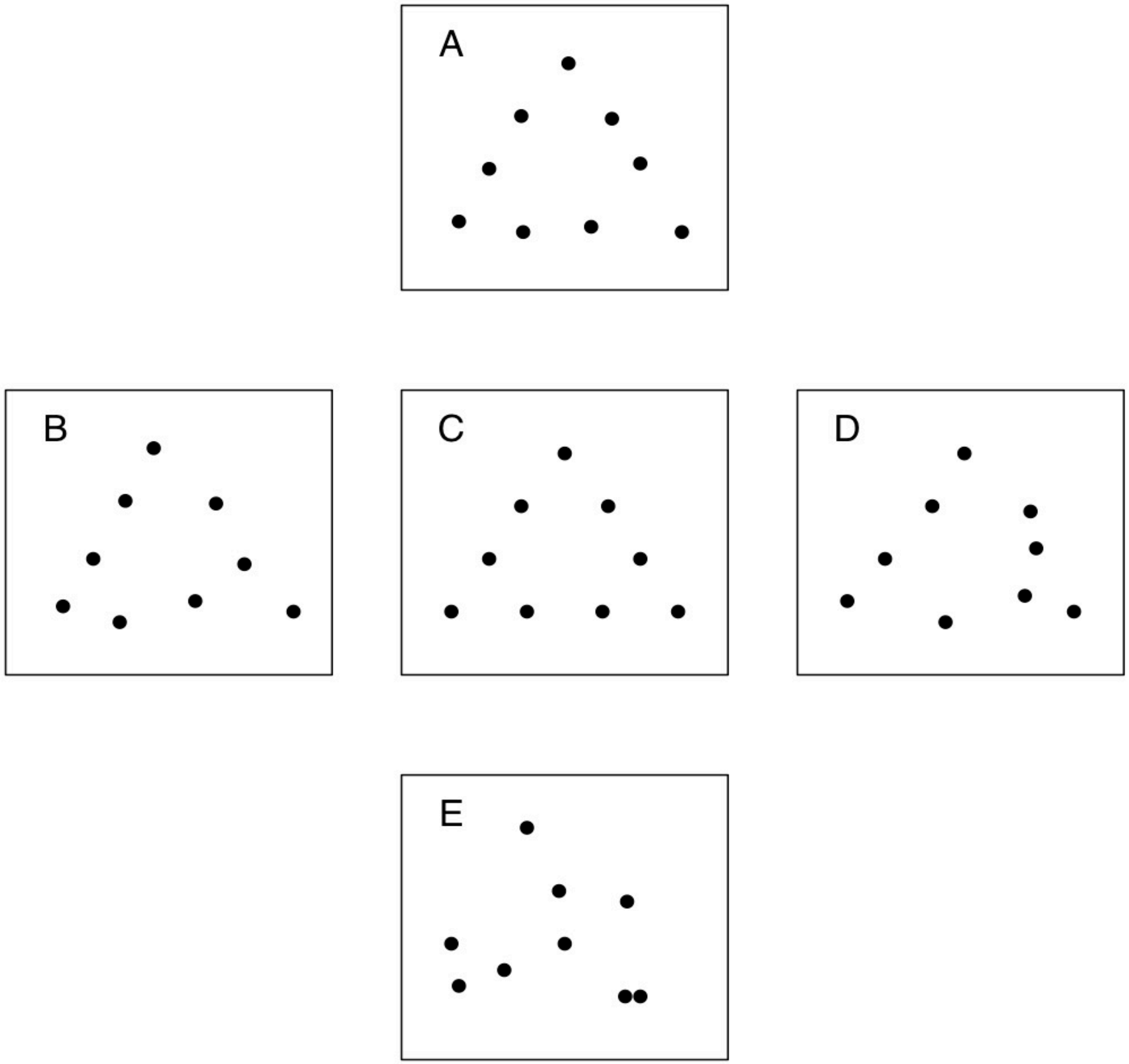
Categorization as typicality

How can we explain typicality?

- One answer: reject definitions, and have a new representation for categories

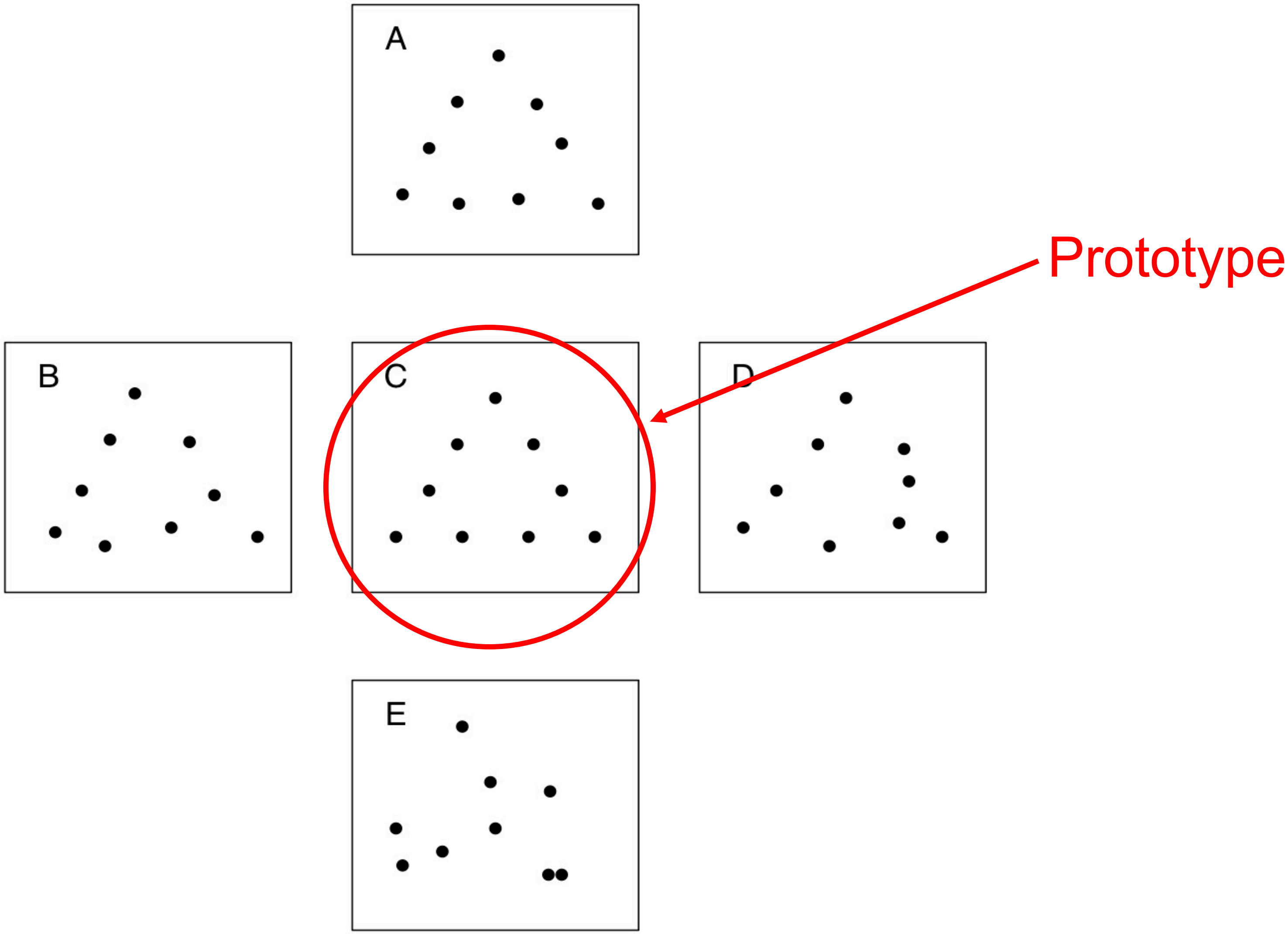
- Prototype theory:
  - categories are represented by a prototype
  - members share a family resemblance to the prototype
  - typicality is a function of similarity to the prototype

Prototypes



(after Posner & Keele, 1968)

Prototypes



(after Posner & Keele, 1968)

Formalizing prototype theories

- **Representation:**
  - Each category (e.g., *A*) has a corresponding prototype (e.g., $\square_A$)

- **Categorization:**
  - Choose category that minimizes (maximizes) the distance (similarity) from *x* to its prototype (Reed, 1972)

Formalizing prototype theories

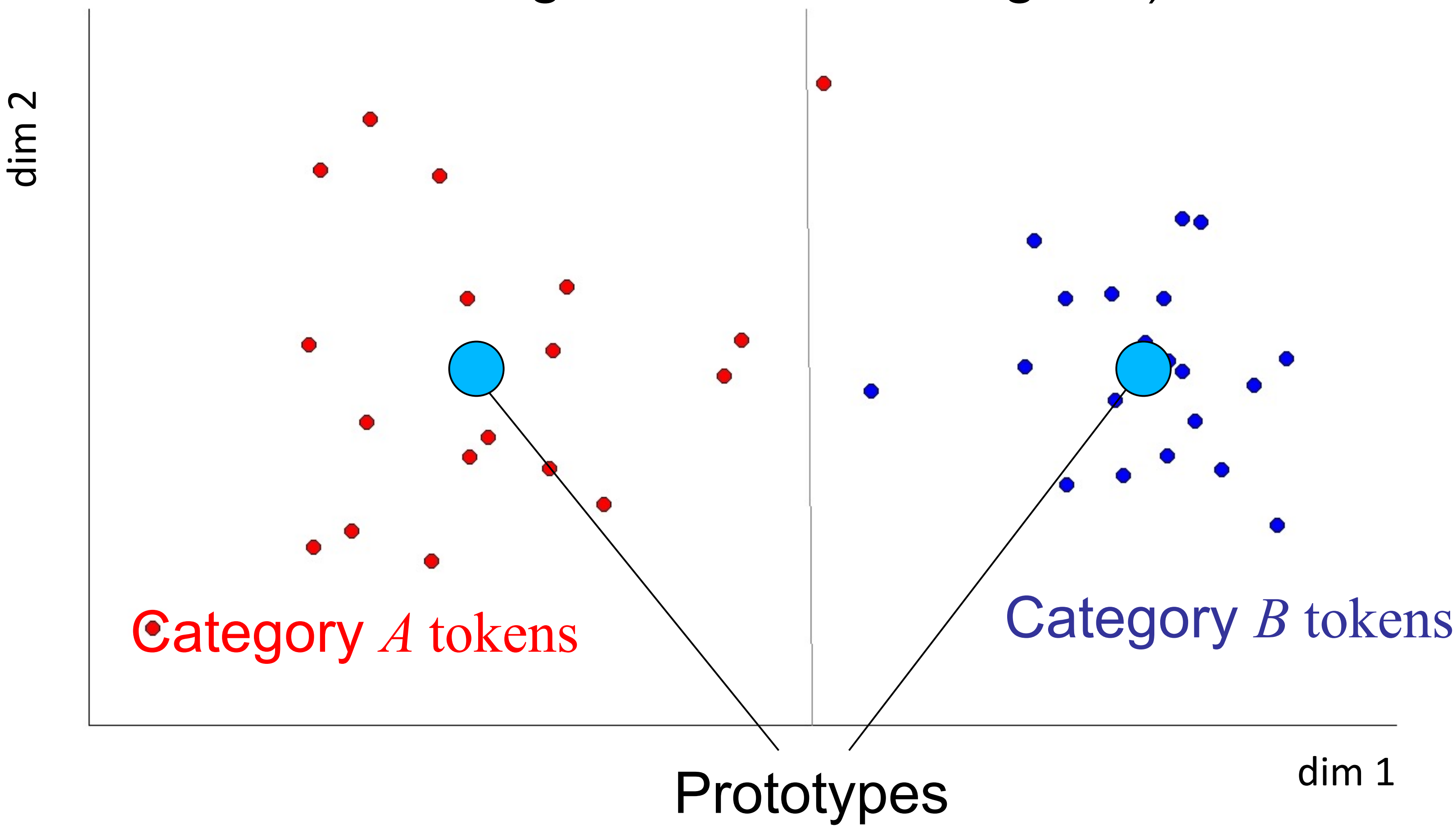
- Single thing stored per category
    - Often: most frequent or "typical" member
- **Representation**: e.g. average of all inputs so far
- **Distance:** e.g. Euclidean

$$d(x, A) = \left[\sum_i (x_i - \eta_{A,i})^2 \right]^{1/2}$$

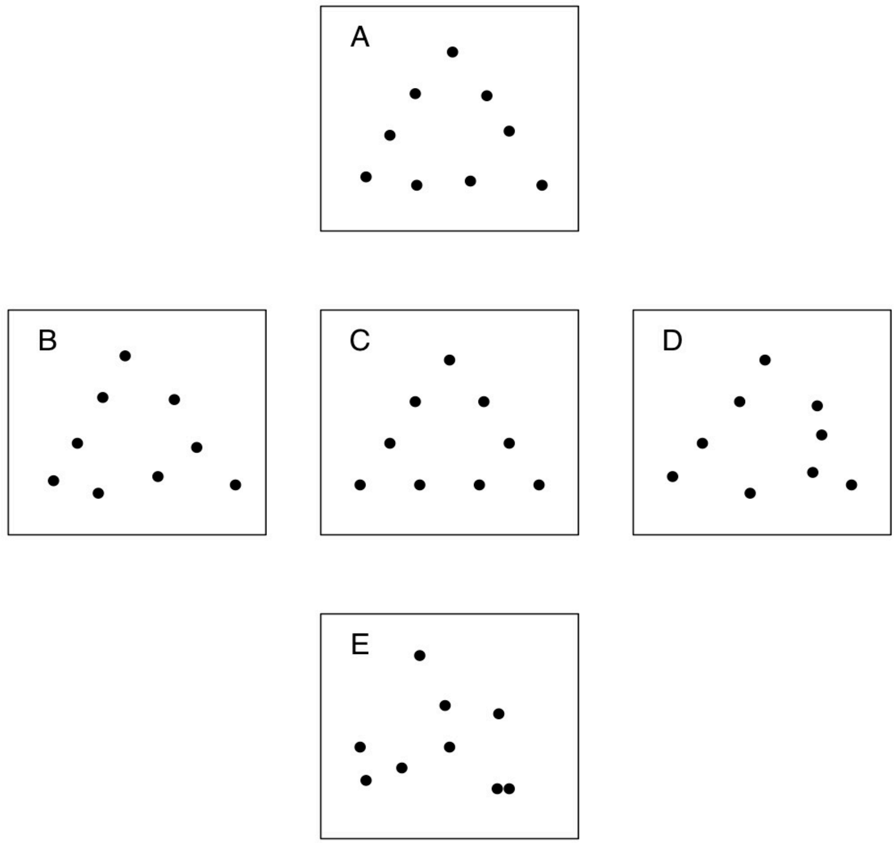
- de Boer: Weighted Euclidean in (F1, F2') space

Formalizing prototype theories

Decision boundary at equal distance (always a straight line for two categories)



Exemplar theories



Store every member ("exemplar") of the category

Formalizing exemplar theories

- Multiple things stored per category
  - Pierrehumbert: All members
  - Wedel: *n* most recent members

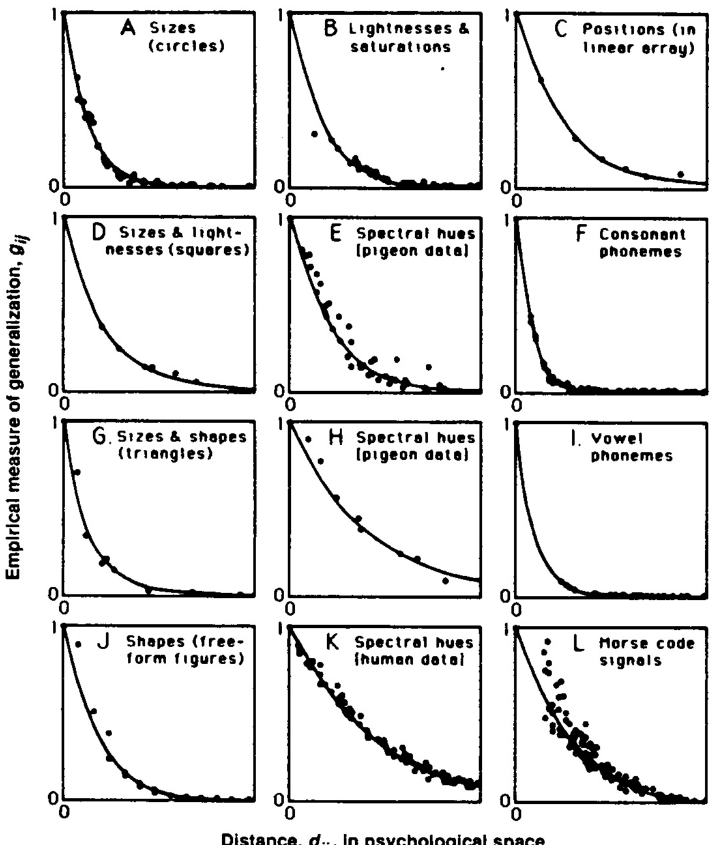
- **Representation:** e.g. a set of stored exemplars $y_1, y_2, \ldots, y_n$, each with a weight

- **Distance**: Window function, or GCM (next slide)

The Generalized Context Model

(Nosofsky, 1986)

Defined for stimuli in psychological space



$$\eta_{ij} = \sum_j \exp(-cd_{ij})^p$$

where

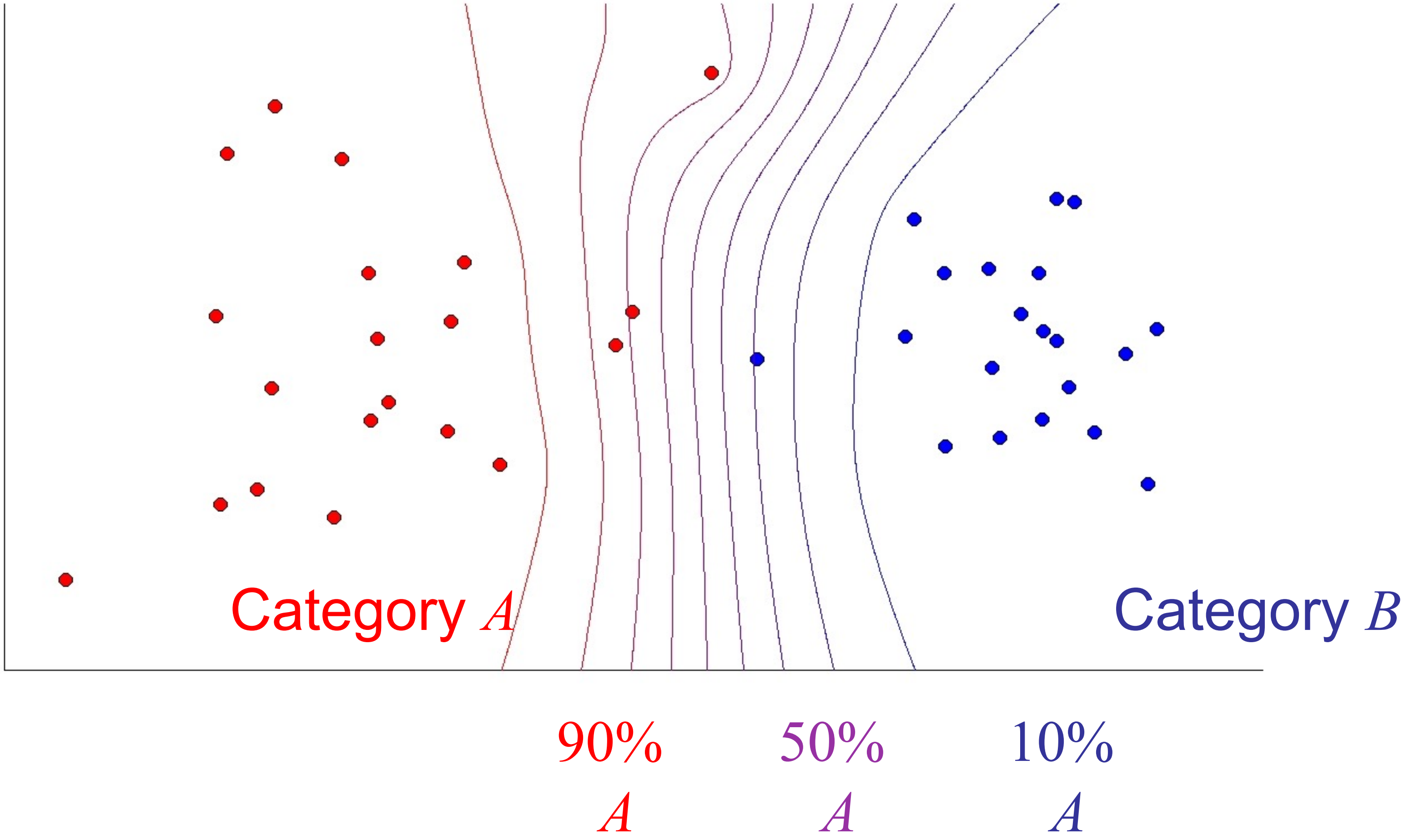
c is "sensitivity"

p = 1 is exponential

p = 2 is Gaussian

The generalized context model

Decision boundary determined by exemplars



Luce-Shepard choice rule

- Common categorization rule
- Choose category A with probability

$$P(A|x) = \frac{\beta_A \sum_{y \in A} \eta_{xy}}{\beta_A \sum_{y \in A} \eta_{xy} + \beta_B \sum_{y \in B} \eta_{xy}}$$

$\square_{xy}$ is similarity of *x* to *y*

$\square_{A}$ is bias towards *A*

Distance function ≠ **decision** rule

- Determining 'distance' between *a* and *b* doesn't tell you what to do with that distance
  - Distance between probe and prototype + stochastic decision rule
  - Exemplar-weighted similarity + deterministic decision rule

- P & deBoer: deterministic

Prototypes vs. exemplars

- Both approaches seem reasonable...?
- Both have been applied in speech
  (prototypes: Samuel 1982, Kuhl 1991; exemplars: Pisoni 1992, etc.)
- In two weeks: categorization as density estimation

Background: Vowel inventory typology

- UPSID: 451 inventories
  (Maddieson, 1984; Maddieson & Precoda, 1990)

- Certain regions of formant space preferred (e.g. point vowels)

- Central vowels
  - largely only introduced at n=6, n=9+
  - central vowels ⇒ peripheral vowels

- n=5-7 : most common
- n=1-2 : vertical

- Secondary articulations
  - Most common: Length, nasality
  - The norm for n=9+

Explanations for patterns

- Structural ("innate")
  (Clements, 2003; Dresher, 2003; Rice, 1999; Government Phonology…)
  – Preferred feature, segment values
  – Preferred feature, segment inventories
    (e.g. symmetry, contrastive hierarchy..)

- Functional
  – Quantal Theory (Stevens 1972, 1989)
  – Versions of dispersion
    (Liljencrants & Lindblom, 1972; Flemming, 1995 et seq; many others)
  – Combo (e.g. Carré 1994)

Dispersion

- Idea: Vowels are optimally dispersed in some space
  - Usually: Acoustic
- Liljencrants & Lindblom (1972)
  - (Day 1)

de Boer (1999, 2000)

- Previous work:
  1. Assumes innate structures, or biases
  2. Global optimization of inventories
  3. Considers inventories of individuals

- Questions:
  - How does population-level inventory result from individual actions/learning?
  - Can we get self-organizing behavior without assuming (1) or (2)?

Model assumptions: High level

- No <u>global</u> optimization
  (c.f. Flemming, Lijencrants & Lindblom)

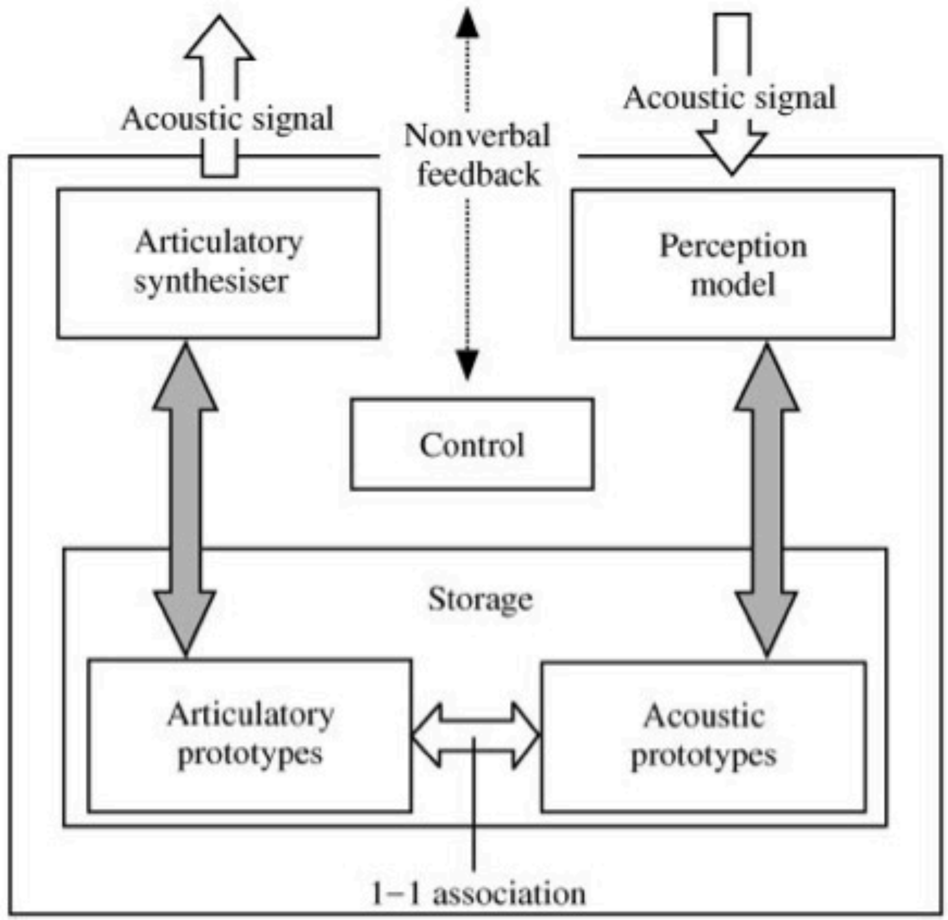
- Update based on imitations only
  (no talking to self)

- No "looking inside partner's head"
  - Don't get partner's <u>intended category</u>
    (c.f. Boersma & Hamann, Wedel)

Goals, non-goals

- Not modeling acquisition
  - Children learn vowel systems differently
- Not modeling historical evolution
  - “Many more complex mechanisms play a role in historical language change than can be modeled by a simple computer simulation.” (454)
- “Although it is not claimed that the model presented here is an accurate model of how humans learn vowel systems, it probably does capture the most important aspects of this process. It is therefore a good model for testing whether self-organization can explain the universal tendencies of human vowel inventories.” (453)

Agent architecture



Articulatory representation

- [height, back, round] $\in [0,1]^3$

- No noise in production

Vowel	*p*	*h*	*r*
[a]	0	0	0
[ɶ]*	0	0	1
[ɐ]	0.5	0	0
[ɐ̹]*	0.5	0	1
[ɑ]	1	0	0
[ɒ]	1	0	1
[e]	0	0.5	0
[ø]	0	0.5	1
[ə]	0.5	0.5	0
[e]	0.5	0.5	1
[ɤ]	1	0.5	0
[o]	1	0.5	1
[i]	0	1	0
[y]	0	1	1
[ɨ]	0.5	1	0
[ʉ]	0.5	1	1
[ɯ]	1	1	0
[u]	1	1	1

Articulatory synthesizer

- Step 1: Articulation → F1-F4

$$\begin{aligned}F_1 = &((-392 + 392r)h^2 + (596 - 668r)h + (-146 + 166r))p^2 \\ &+ ((348 - 348r)h^2 + (-494 + 606r)h + (141 - 175r))p \\ &+ ((340 - 72r)h^2 + (-796 + 108r)h + (708 - 38r))\end{aligned}$$

$$\begin{aligned}F_2 = &((-1200 + 1208r)h^2 + (1320 - 1328r)h + (118 - 158r))p^2 \\ &+ ((1864 - 1488r)h^2 + (-2644 + 1510r)h + (-561 + 221r))p \\ &+ ((-670 + 490r)h^2 + (1355 - 697r)h + (1517 - 117r))\end{aligned}$$

…

- Step 2: Add noise

drawn uniformly from $[-\frac{\psi_{ac}}{2}, \frac{\psi_{ac}}{2}]$

$$F_i = F_i(1 + v_i)$$

Articulatory synthesizer

- Step 1: Articulation → F1-F4

$$\begin{aligned} F_1 = & ((-392 + 392r)h^2 + (596 - 668r)h + (-146 + 166r))p^2 \\ & + ((348 - 348r)h^2 + (-494 + 606r)h + (141 - 175r))p \\ & + ((340 - 72r)h^2 + (-796 + 108r)h + (708 - 38r)) \end{aligned}$$

$$\begin{aligned} F_2 = & ((-1200 + 1208r)h^2 + (1320 - 1328r)h + (118 - 158r))p^2 \\ & + ((1864 - 1488r)h^2 + (-2644 + 1510r)h + (-561 + 221r))p \\ & + ((-670 + 490r)h^2 + (1355 - 697r)h + (1517 - 117r)) \end{aligned}$$

…

- Step 2: Add noise

$$[-\frac{\psi_{ac}}{2}, \frac{\psi_{ac}}{2}]$$ parameter

$$F_i = F_i(1 + v_i)$$

Prototypes

- Vowel *v* stored as prototype
  1. Articulatory aspect: [h, b, r]
  2. Acoustic aspect: F1-F4, synthesized

- Also:
  - Success count: s_v
  - Use count: u_v

Prototypes

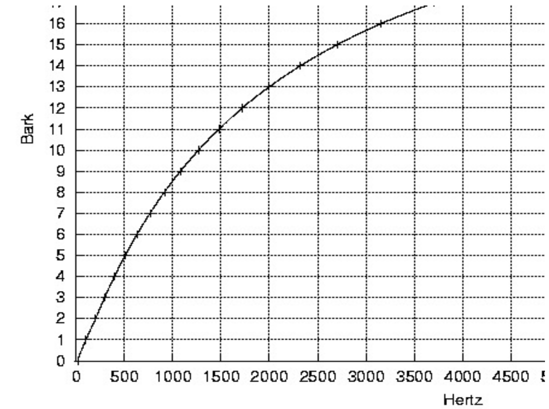
- Vowel *v* stored as prototype
  1. Articulatory aspect: [h, b, r]
  2. Acoustic aspect: F1-F4, synthesized

- Also:
  - Success count: s_v
  - Use count: u_v

- Note: Exemplars not stored

Perception

1. (F1, F2, F3, F4): Hz → bark



1. (F1, F2, F3, F4) → (F1, F2’)
   - F2’: Effective second-formant frequency
     (Mantakas, Schwartz, Escudier, 1986)
   - Roughly: Weight formant peaks closer than critical distance *c* (3.5 bark)

- Formants of higher frequency are “blurred”

$$F'_2 = \begin{cases} F_2 & \text{if } F_3 - F_2 > c \\ \dfrac{(2 - w_1)F_2 + w_1 F_3}{2} & \text{if } F_3 - F_2 \leq c \quad \text{and} \quad F_4 - F_2 > c \\ \dfrac{w_2 F_2 + (2 - w_2)F_3}{2} - 1 & \text{if } F_4 - F_2 \leq c \quad \text{and} \quad F_3 - F_2 < F_4 - F_3 \\ \dfrac{(2 + w_2)F_3 - w_2 F_4}{2} - 1 & \text{if } F_4 - F_2 \leq c \quad \text{and} \quad F_3 - F_2 \geq F_4 - F_3 \end{cases}$$

$$w_1 = \frac{c - (F_3 - F_2)}{c} \qquad\qquad w_2 = \frac{(F_4 - F_3) - (F_3 - F_2)}{F_4 - F_2}$$

$$F_2' = \begin{cases} F_2 & \text{if } F_3 - F_2 > c \\ \dfrac{(2-w_1)F_2 + w_1 F_3}{2} & \text{if } F_3 - F_2 \le c \quad \text{and} \quad F_4 - F_2 > c \\ \dfrac{w_2 F_2 + (2-w_2)F_3}{2} - 1 & \text{if } F_4 - F_2 \le c \quad \text{and} \quad F_3 - F_2 < F_4 - F_3 \\ \dfrac{(2+w_2)F_3 - w_2 F_4}{2} - 1 & \text{if } F_4 - F_2 \le c \quad \text{and} \quad F_3 - F_2 \ge F_4 - F_3 \end{cases}$$

$$w_1 = \frac{c - (F_3 - F_2)}{c} \qquad w_2 = \frac{(F_4 - F_3) - (F_3 - F_2)}{F_4 - F_2}$$

"a rather ad-hoc solution" (449), because strength of formant peaks not calculated

Perceptual distance

$$D = \sqrt{(F_1^a - F_1^b)^2 + \lambda(F_2^{a'} - F_2^{b'})^2}$$

relative weight of higher formants
w.r.t. F1

- de Boer: λ= 0.3 (cites previous work)
- Interpretation: How much vowel space "stretched" horizontally vs. vertically

Model details: Imitation game

- *nAgents* : # of agents
- *nIts* : # of interactions

1. initialize agents with empty inventories

2. do *nIts* interactions:
   - pick two agents at random
   - agents interact

TABLE III. Basic organization of the imitation game

	Imitator	Imitator	

1	If (V = Ø)
	  Add random vowel to V
	Pick random v vowel v from V
	$u_v \leftarrow u_v + 1$
	Produce signal A_1:
	  $A_1 \leftarrow ac_v + noise$		

		Receive signal A_1
		If ($V = \emptyset$)
		  Find phoneme (v_{news}, A_1)
		  $V \leftarrow V \cup v_{new}$
		Cacalculate v_{rec}:
		$v_{rec} \in V \wedge \neg\exists v_2 : (v_2 \in V \wedge D(A_1, ac_{v2}) < D(A_1, ac_{vrec}))$
		Produce signal A_2:
		$A_2 \leftarrow ac_{vrec} + noise$
			2

3	Receive signal A_2.
	Calculate v_{rec}:
	$v_{rec} \in V \wedge \neg\exists v_2 : (v_2 \in V \wedge D(A_2, ac_{v2}) < D(A_2, ac_{vrec}))$
	If ($v_{rec} = v$)
	  Send nonverbal feedback:
	    success.
	  $s_v \leftarrow s_v + 1$
	Else
	  Send nonverbal feedback:
	    failure.		

		Receive nonverbal feedback **Update V according to feedback signal.**	4
5	**Do other updates of V.**	**Do other updates of V.**	5

Step 1: Agent 1

Step 2: Agent 2

Step 3: Agent 1

Step 4: Agent 2

Step 5: Agents 1 & 2

Shift closer (v, A)	**Find phoneme** (v_{new}, A)	**Update according to feedback signal**
$v_{best} \leftarrow v$	$ar_v \leftarrow (0.5, 0.5, 0.5)$	$v_{vrec} \leftarrow u_{vrec} + 1$
For (all six neighbors v_{neigh} of v) do:	$ac_v \leftarrow S(ar_v)$	If (feedback signal $=$ *success*)
If $(D(ac_{vneigh}, A) <$	$s_v \leftarrow 0$	**Shift closer** (v_{rec}, A_1)
$D(ac_{vrec}, A)$		
$v_{best} \leftarrow v_{neigh}$	$u_v \leftarrow 0$	$s_{vrec} \leftarrow s_{vrec} + 1$
$v \leftarrow v_{best}$	Do	Else
	$v_{new} \leftarrow v$	If $(u_{vrec}/s_{vrec} > threshold)$
	Shift closer (v_{new}, A)	**Find phoneme**
		(v_{new}, A_1)
	Until $(v = v_{new})$	$V \leftarrow V \cup v_{new}$
		Else
		Shift closer (v_{rec}, A_1)

ε_{artic} away, 6 directions

Shift closer (v, A)	**Find phoneme** (v_{new}, A)	**Update according to feedback signal**
$v_{best} \leftarrow v$	$ar_v \leftarrow (0.5, 0.5, 0.5)$	$v_{vrec} \leftarrow u_{vrec} + 1$
For (all six neighbors v_{neigh} of v) do:	$ac_v \leftarrow S(ar_v)$	If (feedback signal $=$ *success*)
If $(D(ac_{vneigh}, A) <$	$s_v \leftarrow 0$	**Shift closer** (v_{rec}, A_1)
$D(ac_{vrec}, A))$		
$v_{best} \leftarrow v_{neigh}$	$u_v \leftarrow 0$	$s_{vrec} \leftarrow s_{vrec} + 1$
$v \leftarrow v_{best}$	Do	Else
	$v_{new} \leftarrow v$	If $(u_{vrec}/s_{vrec} > threshold)$
	Shift closer (v_{new}, A)	**Find phoneme**
		(v_{new}, A_1)
	Until $(v = v_{new})$	$V \leftarrow V \cup v_{new}$
		Else
		Shift closer (v_{rec}, A_1)

0.5

Step 5

Merge (v_1, v_2, V)
If $(s_{v1}/u_{v1} < s_{v2}/u_{v2})$
  $s_{v2} \leftarrow s_{v2} + s_{v1}$
  $u_{v2} \leftarrow u_{v2} + u_{v1}$
  $V \leftarrow V - v_1$
Else
  $s_{v1} \leftarrow s_{v1} + s_{v2}$
  $\mathrm{u}_{v1} \leftarrow \mathrm{u}_{v1} + \mathrm{u}_{v2}$
$V \leftarrow V - v_2$

Do other updates of V
For $(\forall v \in V)$//Remove bad vowels
  if ($s_v/u_v <$ *throwaway threshold* $\wedge\, u_v >$ *min. uses*)
    $V \leftarrow V - v$
For $(\forall v_1 \in V)$//Merging of vowels
  For $(\forall v_2 : (v_2 \in V \wedge v_2 \neq v_1))$
    If ($D(ac_{v1}, ac_{v2}) <$ *acoustic merge threshold*)
      Merge (v_1, v_2, V)
    If (Euclidean distance between av_{v1} and $av_{v2} <$ *articulatory merge threshold*)
      Merge(v_1, v_2, V)
Add new vowel to V with small probability.

Step 5

0.7 5

Merge (v_1, v_2, V)
If $(s_{v1}/u_{v1} < s_{v2}/u_{v2})$
  $s_{v2} \leftarrow s_{v2} + s_{v1}$
  $u_{v2} \leftarrow u_{v2} + u_{v1}$
  $V \leftarrow V - v_1$
Else
  $s_{v1} \leftarrow s_{v1} + s_{v2}$
  $u_{v1} \leftarrow u_{v1} + u_{v2}$
$V \leftarrow V - v_2$

Do other updates of V
For $(\forall v \in V)$//Remove bad vowels
  if $(s_v/u_v <$ *throwaway threshold* $\wedge\ u_v >$ *min. uses*$)$
    $V \leftarrow V - v$
For $(\forall v_1 \in V)$//Merging of vowels ??
  For $(\forall v_2 : (v_2 \in V \wedge v_2 \neq v_1))$ 0.17
    If $(D(ac_{v1}, ac_{v2}) <$ *acoustic merge threshold*$)$
      Merge (v_1, v_2, V)
    If (Euclidean distance between av_{v1} and $av_{v2} <$ *articulatory merge threshold*)
      Merge(v_1, v_2, V)
Add new vowel to V with small probability. 0.01

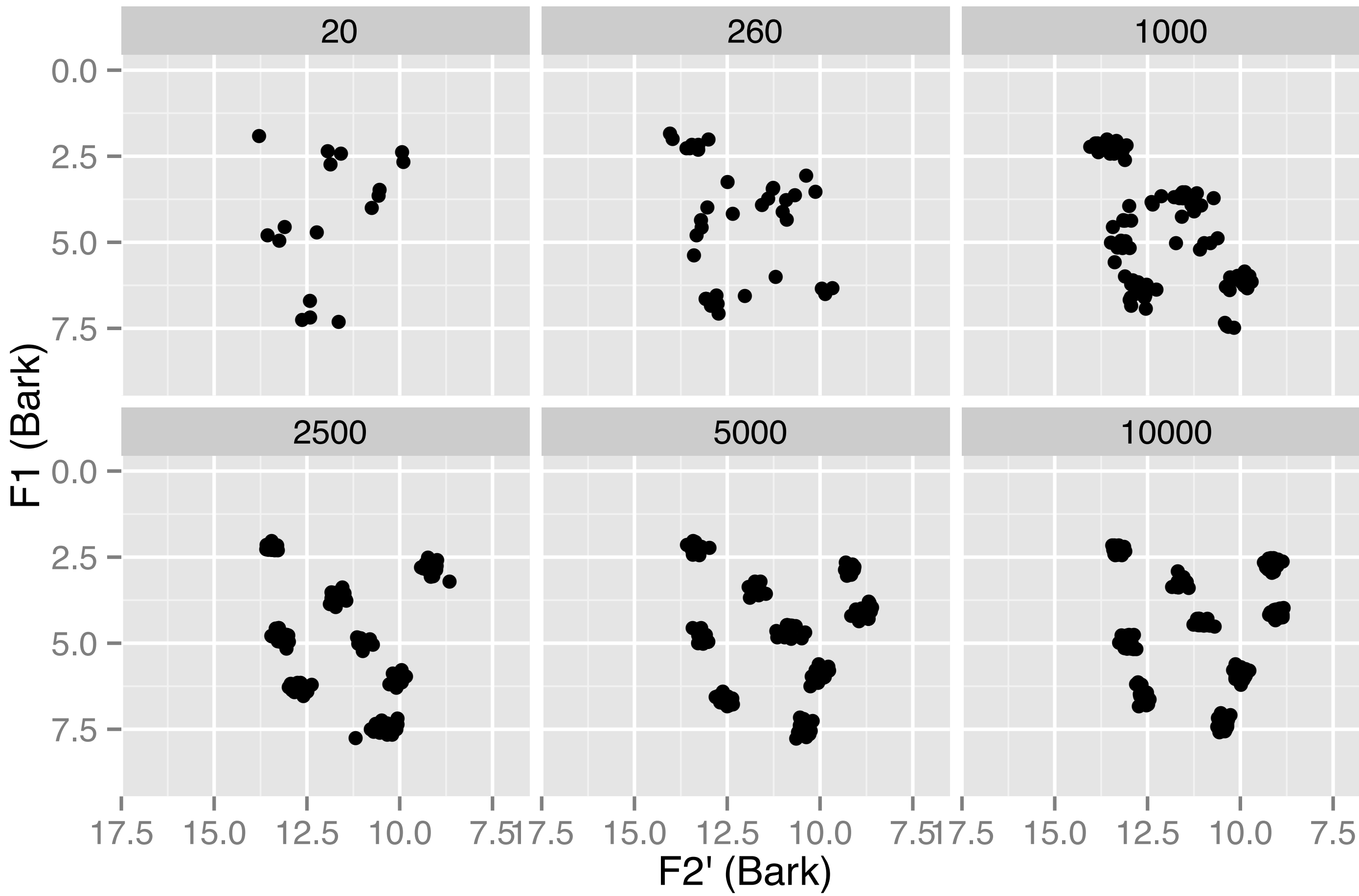
- What do we do with
  - 3 vowels close enough to be merged?
  - vowel with $u_v = 0$?

Parameters

ψ_{ac}	noise
λ	relative weight of F2' and F1
n_{agents}	# agents
n_{its}	# iterations
ε_{artic}	articulatory step size for shifting prototypes
$\vdots$	(other params held constant today)

Sample run

- n_{its}=10000, n_{agents}=20, ψ=0.1, ε_{artic}=0.03, λ= 0.3



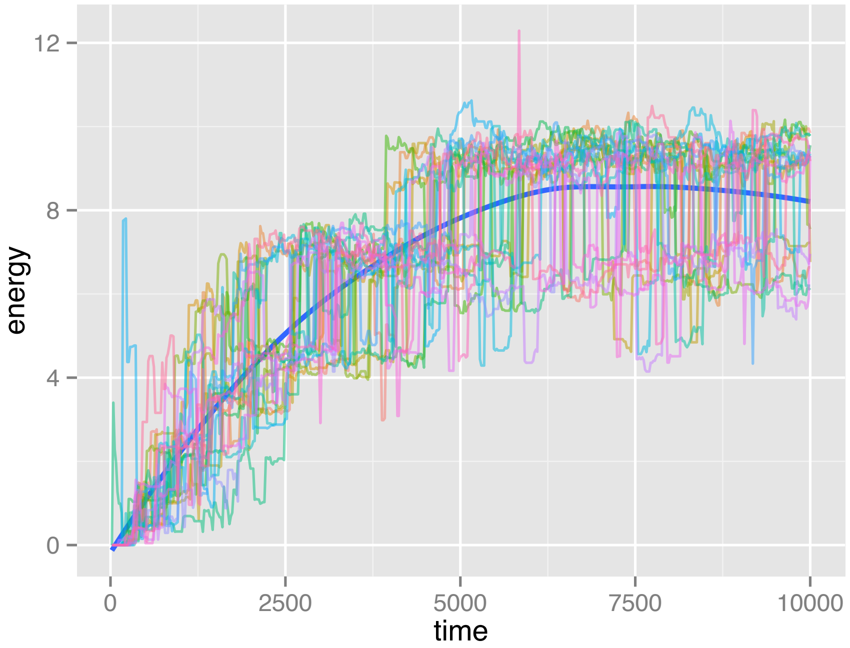
Evaluation measures

- Success rate (1 agent): $SR = \sum_v \frac{s_v}{u_v}$

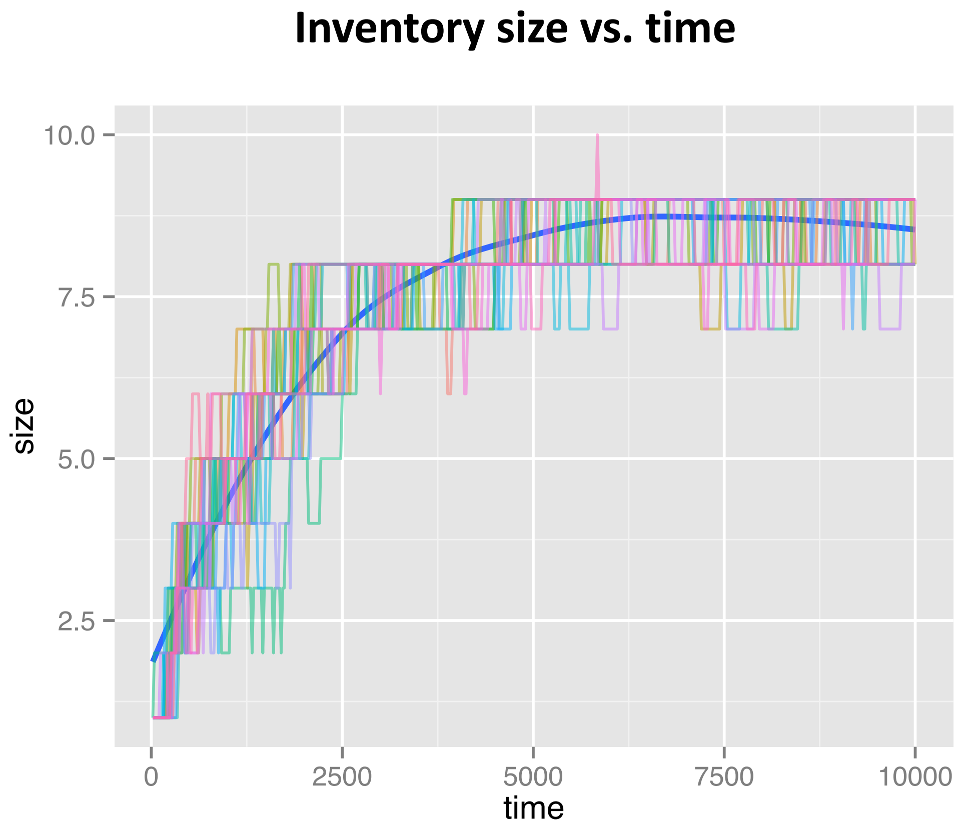
Evaluation measures

- Inventory energy: $$E = \sum_{(v_1,v_2)\,pairs} \frac{1}{D(v_1,v_2)^2}$$
- Motivation:
  "…minimizing the energy function has been shown to result in realistic vowel systems"
  (e.g. Liljencrants & Lindblom, 1972)

Inventory energy vs. time

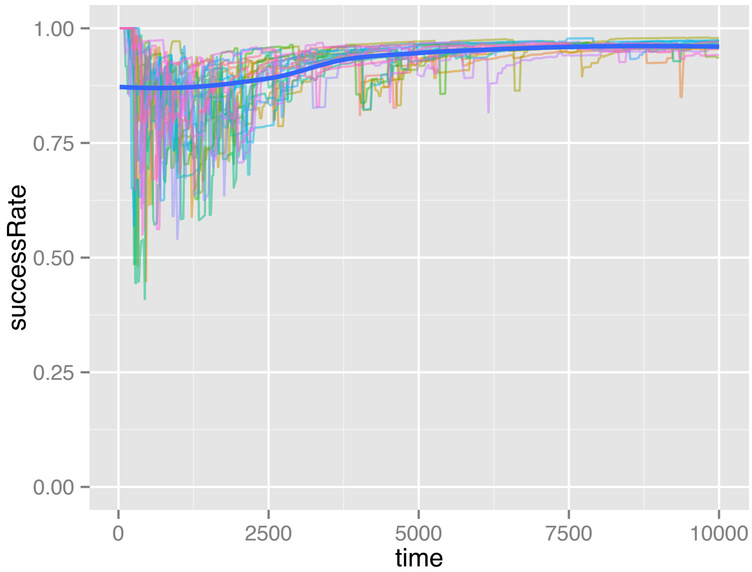


Evaluation measures

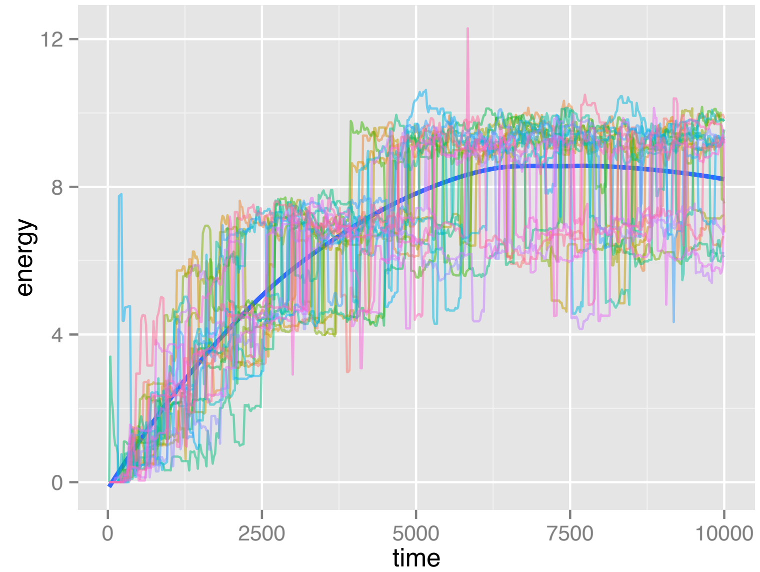


- Success prob, energy, inventory size all stabilize around t=5000

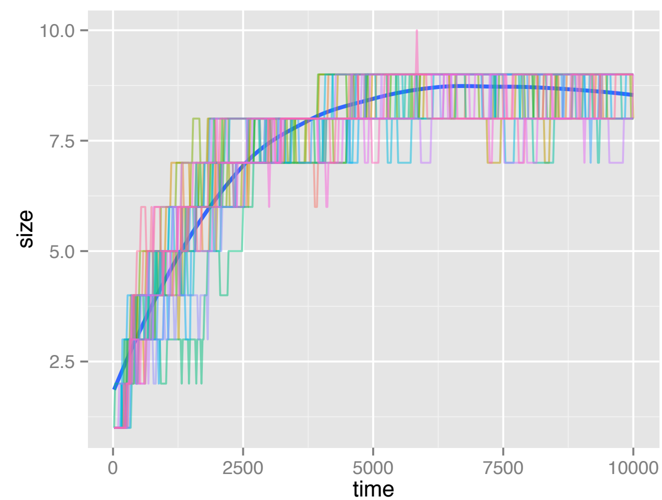
Success rate vs. time



Inventory energy vs. time



Inventory size vs. time

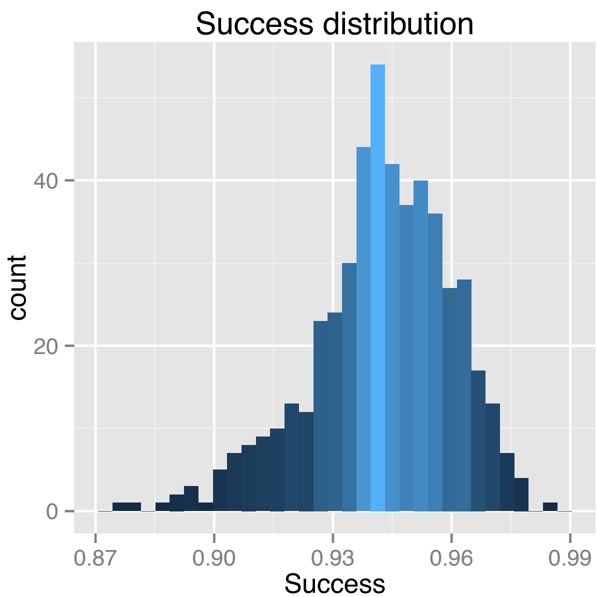


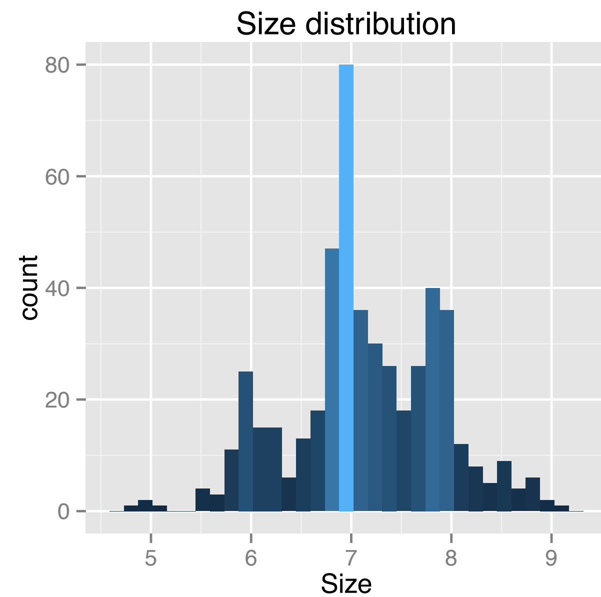
- Why these trajectories?
- Why are energy and size trajectories similar?

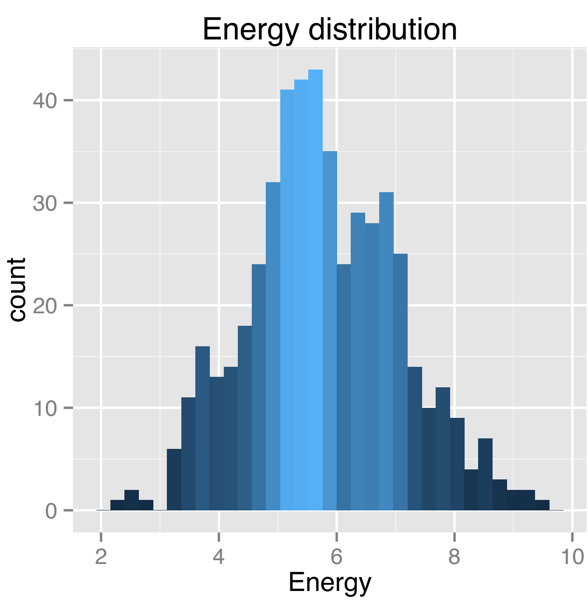
- Finding 1: Vowel systems emerge (for some parameter settings)
  - Individuals converge on similar systems of vowels
  - Relatively stable over time

- How variable is the resulting system, for fixed parameter settings?

1. Run 500 times
   - n_{its}=5000, n_{agents}=20, ψ=0.1, $\varepsilon_{\mathrm{artic}}$=0.03, λ= 0.3
2. Examine histograms of evaluation measures
   - Mean over all agents' inventories after n_{its} interactions



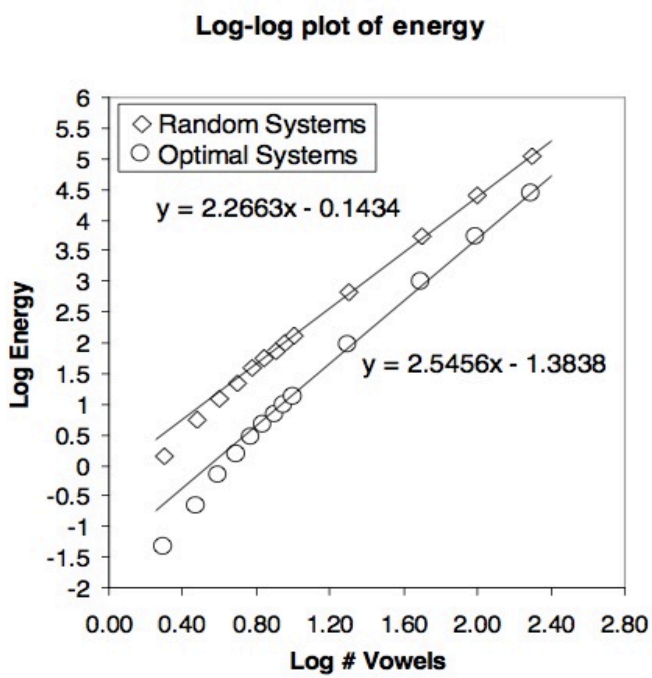




- Size: 7.1 ± 0.8
- Energy: 5.8 ± 1.3
- Success: 0.94 ± 0.02

Random and optimal systems

- Can compare energy, success probability to their distributions for random and optimal systems (de Boer, 1999)

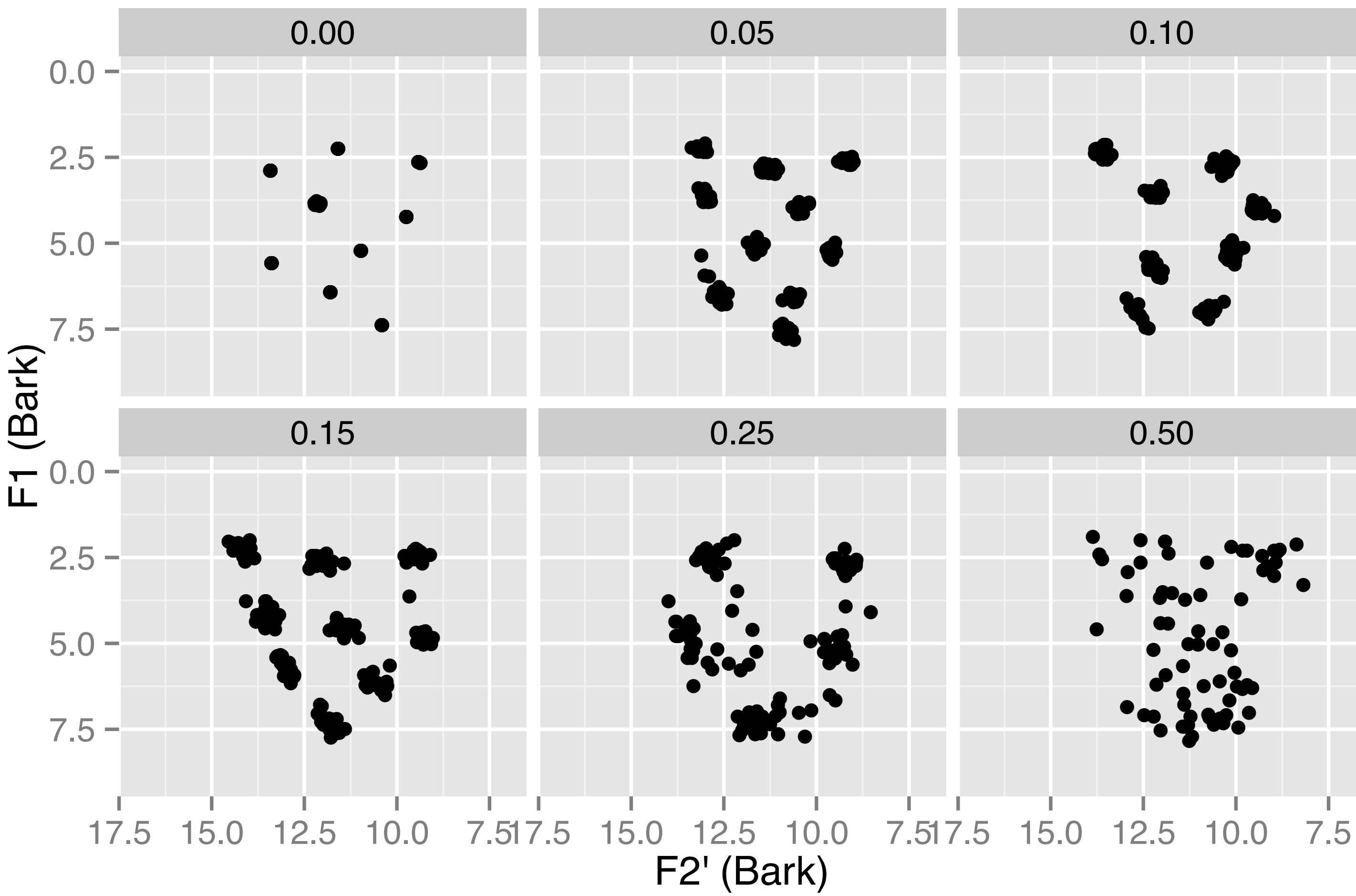


- For constant # of vowels, emerged systems tend to be near-optimal (!)

Varying parameters

- “The question remains whether these results are due to fine-tuning of parameters or whether they are an inevitable result of the interactions between the agents.” (456)

Varying the amount of noise



ψ = 0, 0.05, 0.10, 0.15, 0.25, 0.50

- Increased ψ:
  - Success rate: Decreases (0.98-0.74)
  - Number of vowels: Decreases (9.00-3.60)
  - System coherence across agents: Decreases
  - Energy: Decreases

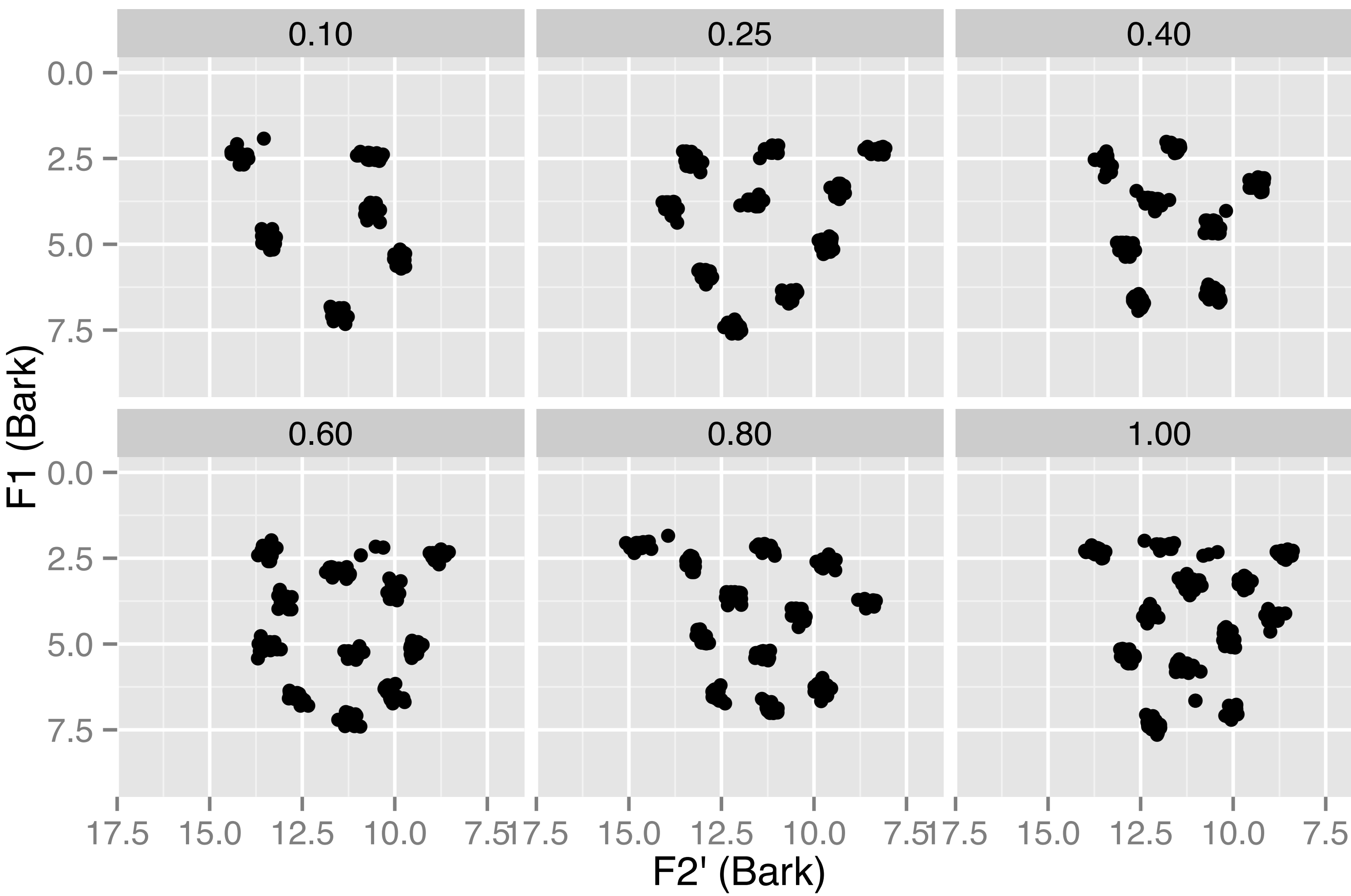
- Why?
- Which ones look “realistic”?

- Increased ψ:
  - Success rate: Decreases (0.98-0.74)
  - Number of vowels: Decreases (9.00-3.60)
  - System coherence across agents: Decreases
  - Energy: Decreases

- Why?
- Which ones look "realistic"?

- Opposite success rate pattern from de Boer
  - Interaction with another parameter we've set differently? (Or a bug?)

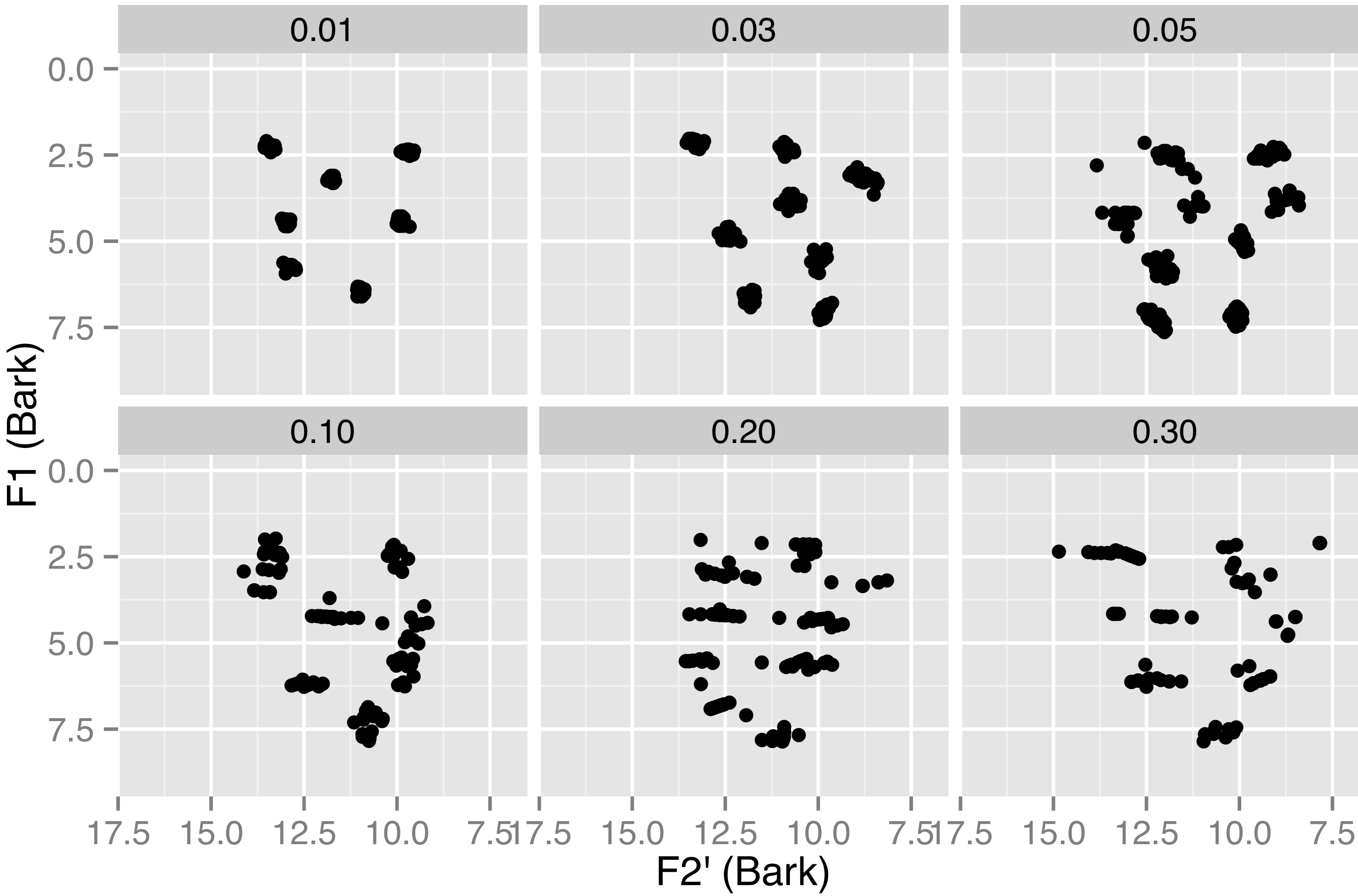
Varying λ


0.10
0.25
0.40
0.60
0.80
1.00
0.0
2.5
5.0
7.5
F1 (Bark)
17.5
15.0
12.5
10.0
7.5
F2' (Bark)

- Increased λ:
  - Success rate: approx. constant
  - Size, energy: Increases
  - More vowel height dimensions, sort of more backness dimensions

- Why?
- Which ones look realistic?

Varying ε_{artic}



- Increased ε_{artic} :
  - Success rate: decreases
  - Energy, size: approx. constant

- Where's the horizontal patterning coming from?

- You know the drill..

Varying n_{agents}

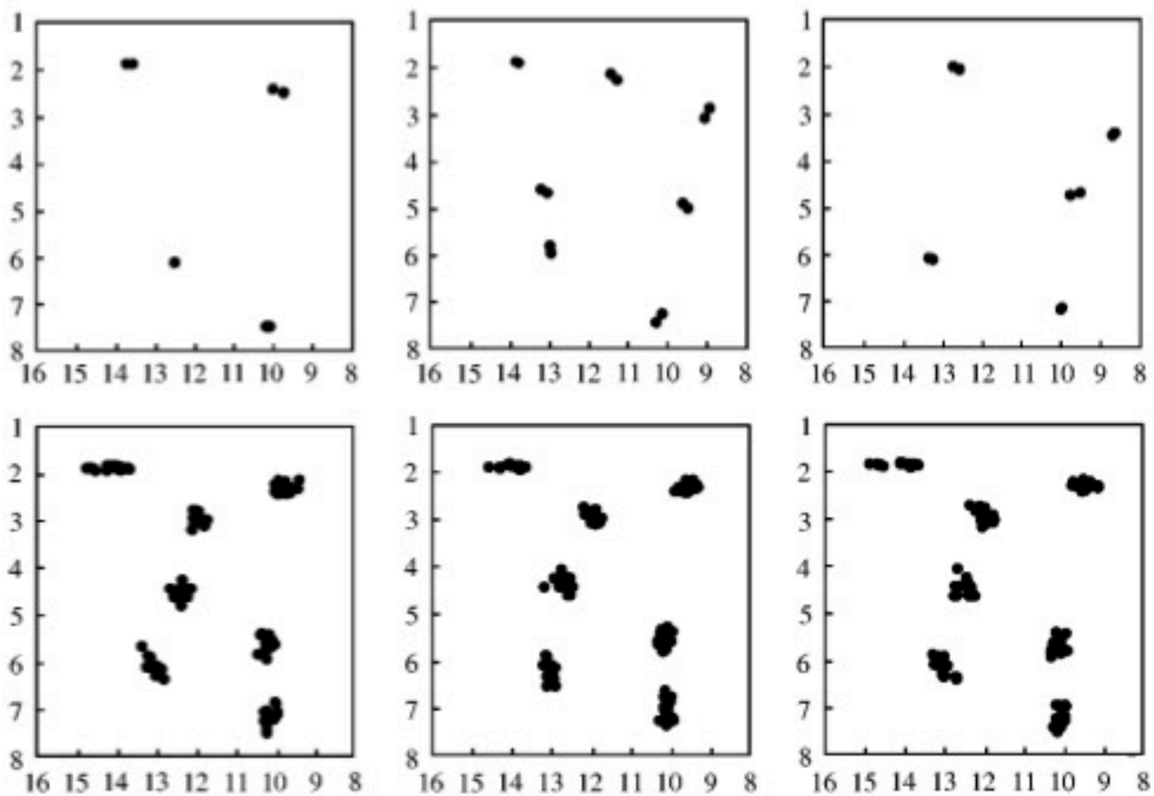
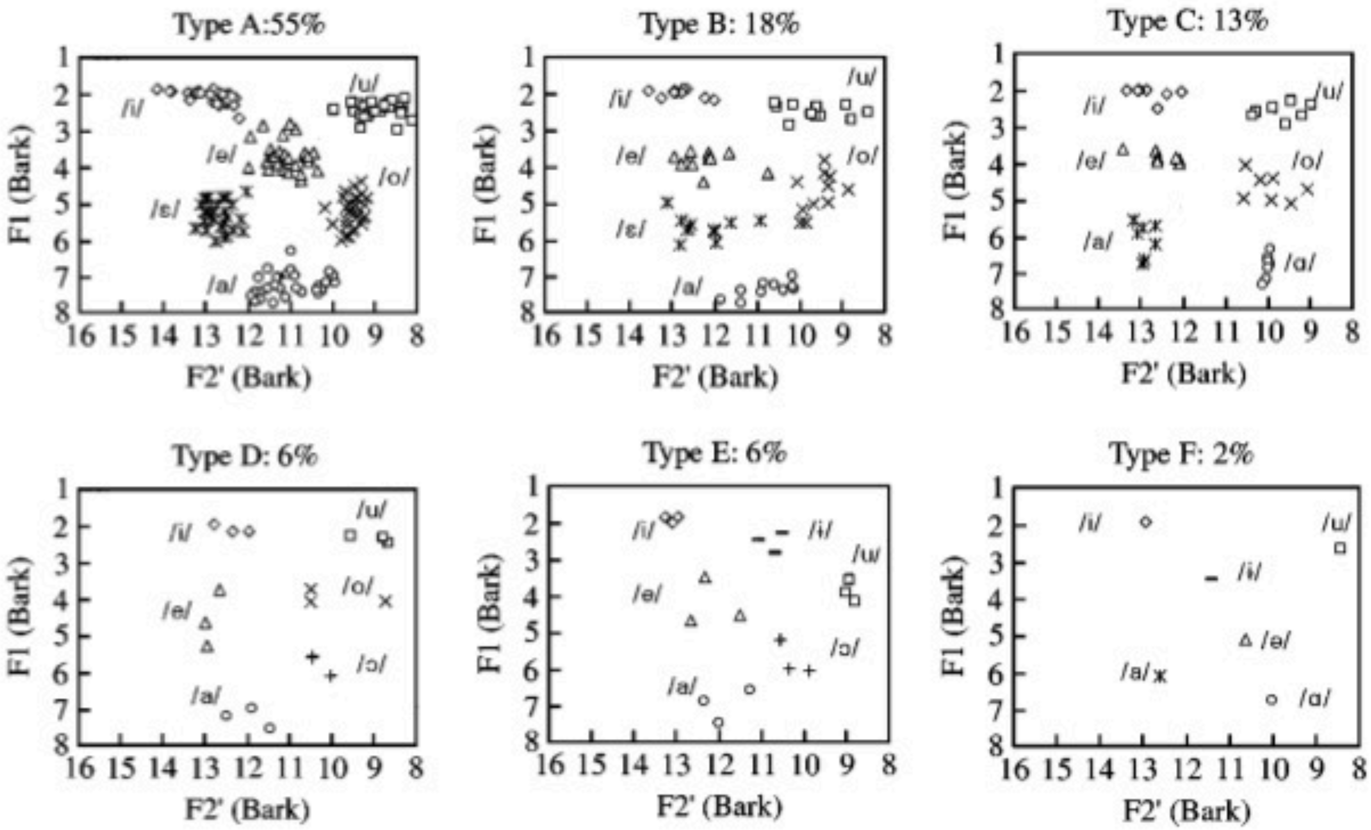


Figure 9. Populations of 2 (top) and 20 (bottom) agents shown at 250, 350 and 450 games per agent from left to right (F'_2 in Bark on x-axis and F_1 in Bark on y-axis). Note the relative instability over time of the vowel system in the small population.

- Increased n_{agents}:
  - Success rate: approx. constant
  - Size, energy: increase (a little)
  - Agents converge, but prototypes move around.

- Why?

Comparisons to empirical data



- Empirical: 43% Type A, 20% Type B, 7% Type C
- Sensitivity to parameters?

- Other predictions within same n:
  - n=3 : Predicts too many “vertical” systems
  - n=4, 5, 7: Very good
  - n=8+ : Worse (c.f. Liljencrants & Lindblom)

- Predictions of relative frequency of different n:
  - Not so clear how to calculate
  - Under one method (Fig 11), reasonable fit, but lower mean N predicted than observed.

What have we learned?

- Existence proofs?
- Explicitness / implementation?
- Counterintuitive results?
- Qualitative predictions?
- Baseline?

Extensions / projects

- Interaction of $\psi, \varepsilon_{artic}$
  - Both influence category size/number

- Varying some (set of) parameters not considered here, or by de Boer
  - acoustic & articulatory merge thresholds, success threshold, mininum number of uses before discarding, random addition probability..

- Prob. of different *n* vowel systems for $n \neq 6$
  - Comparison to empirical data
  - Robustness to parameter changes

- Vary starting state of agents
  - Lots of vowels (vs none; c.f. infants)
  - Canonical systems (like [a, i, e, o, u]): stable states?

- Lifespan: Agents born and die at some rate
  - ε_{artic} changes with age

- Non-random interaction probabilities
  - Two groups; most interaction intra-group
  - Dialect contact

- Incorporate secondary articulations
  - Length, nasality