

Exemplar dynamics and the emergence of categories

Gerhard Jäger

Gerhard.Jaeger@uni-bielefeld.de

January 11, 2007

University of Stuttgart, CRC 732



Overview

- exemplar-based evolution
- Evolutionary Game Theory
- evolutionary stability
- convex meanings
- color terms
- conclusion



Conceptualization of language evolution

prerequisites for evolutionary dynamics

- replication
- variation
- selection



Exemplar dynamics

- empiricist view on language processing/language structure
- popular in functional linguistics (esp. phonology and morphology) and in computational linguistics (aka. “memory-based”)

Basic idea

- large amounts of previously encountered instances (“exemplars”) of linguistic structures are stored in memory
- very detailed representation of exemplars
- little abstract categorization
- similarity metric between exemplars
- new items are processed by analogy to exemplars that are stored in memory



evolutionary exemplar dynamics

- exemplars form populations
- bidirectionality of exemplars and priming lead to replication of exemplars
- replication may be unfaithful \Rightarrow linguistic variation
- differential replication \Rightarrow evolutionary dynamics

How can this dynamics be modeled formally?



Replication

(at least) two modes of exemplar replication:

- acquisition
- priming



Replication

(at least) two modes of exemplar replication:

- acquisition
- priming

Variation

- linguistic creativity
- reanalysis
- language contact
- ...



Replication

(at least) two modes of exemplar replication:

- acquisition
- priming

Variation

- linguistic creativity
- reanalysis
- language contact
- ...

Selection

- social selection
- selection for learnability
- selection for primability



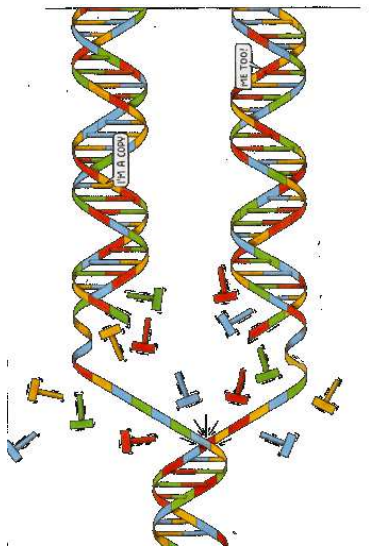
learnability/primability

- selection **against complexity**
- selection **against ambiguity**
- selection **for frequency**



Evolutionary Game Theory

- populations of players
- individuals are (genetically) programmed for certain strategy
- individuals replicate and thereby pass on their strategy



Utility and fitness

- number of offspring is monotonically related to average utility of a player
- high utility in a competition means the outcome improves reproductive chances (and vice versa)
- number of expected offspring (Darwinian fitness) corresponds to **expected utility** against a population of other players
- genes of individuals with high utility will spread



Replicator dynamics

- simplest dynamics that implements these ideas
- fitness is simply identified with utility

$$\frac{dx_i}{dt} = x_i \left(\sum_{j=1}^n y_j u_A(i, j) - \sum_{k=1}^n x_k \sum_{j=1}^n y_j u_A(k, j) \right)$$
$$\frac{dy_i}{dt} = y_i \left(\sum_{j=1}^m x_j u_B(i, j) - \sum_{k=1}^n y_k \sum_{j=1}^m x_j u_B(k, j) \right)$$

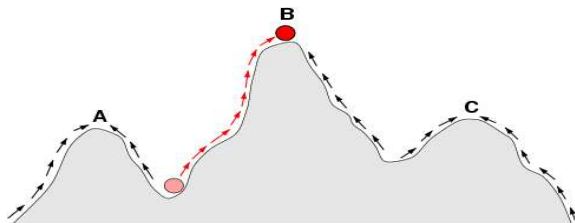
x_i ... proportion of s_i^A within the A -population

y_i ... proportion of s_i^B within the B -population



Evolutionary stability

- Darwinian evolution predicts ascent towards local fitness maximum
- once local maximum is reached: stability
- only random events (genetic drift, external forces) can destroy stability
- central question for evolutionary model: what are stable states?



Evolutionary stability (cont.)

- replication sometimes unfaithful (mutation)
- population is **evolutionarily stable** \leadsto resistant against small amounts of mutation
- Maynard Smith (1982): static characterization of
Evolutionarily Stable Strategies
(ESS) in terms of utilities only



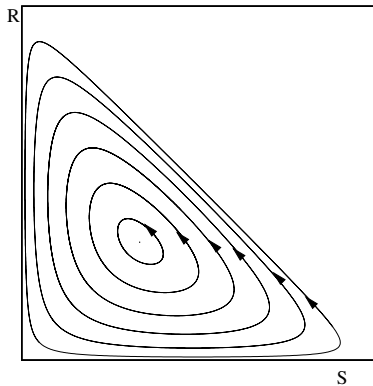
Rock-Paper-Scissor

	R	P	S
R	0	-1	1
P	1	0	-1
S	-1	1	0

- one stationary state (“Nash equilibrium”): $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$
- not evolutionarily stable though



Trajectories



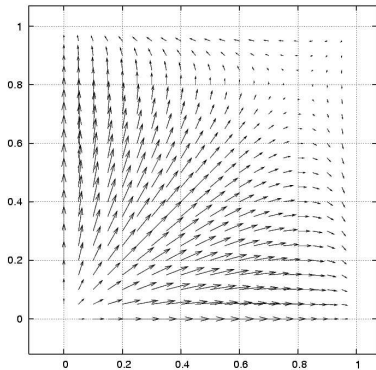
Hawks and Doves

	H	D
H	1,1	7,2
D	2,7	3,3

- two-population setting:
 - both A and B come in hawkish and dovish variant
 - everybody only interacts with individuals from opposite “species”
 - excess of A -hawks helps B -doves and vice versa
 - population push each other into opposite directions



Vector field



Definition (Strict Nash Equilibrium)

A pair of strategies (S, H) is a Strict Nash Equilibrium iff

$$\forall S' (S' \neq S \rightarrow u(S, H) > u(S', H))$$

and

$$\forall H' (H' \neq H \rightarrow u(S, H) > u(S, H'))$$

- in a SNE, S is unique best response to H and vice versa



Evolutionary stability

Definition (Strict Nash Equilibrium)

A pair of strategies (S, H) is a Strict Nash Equilibrium iff

$$\forall S' (S' \neq S \rightarrow u(S, H) > u(S', H))$$

and

$$\forall H' (H' \neq H \rightarrow u(S, H) > u(S, H'))$$

- in a SNE, S is unique best response to H and vice versa

Theorem (Selten 1980)

(S, H) is evolutionarily stable if and only if it is a Strict Nash Equilibrium.



Gärdenfors (2000):

- meanings are arranged in **conceptual spaces**
- conceptual space has geometrical structure
- dimensions are founded in perception/cognition



Gärdenfors (2000):

- meanings are arranged in **conceptual spaces**
- conceptual space has geometrical structure
- dimensions are founded in perception/cognition

Convexity

A subset C of a conceptual space is said to be *convex* if, for all points x and y in C , all points between x and y are also in C .



Gärdenfors (2000):

- meanings are arranged in **conceptual spaces**
- conceptual space has geometrical structure
- dimensions are founded in perception/cognition

Convexity

A subset C of a conceptual space is said to be *convex* if, for all points x and y in C , all points between x and y are also in C .

Criterion P

A *natural property* is a convex region of a domain in a conceptual space.



Examples

- spatial dimensions: *above, below, in front of, behind, left, right, over, under, between ...*
- temporal dimension: *early, late, now, in 2005, after, ...*
- sensual dimensions: *loud, faint, salty, light, dark, ...*
- abstract dimensions: *cheap, expensive, important, ...*



The naming game

- two players:
 - **S**peaker
 - **H**earer
- infinite set of **M**eanings, arranged in a finite metrical space
distance is measured by function $d : M^2 \mapsto R$
- finite set of **F**orms
- sequential game:
 - 1 nature picks out $m \in M$ according to some probability distribution p and reveals m to S
 - 2 S maps m to a form f and reveals f to H
 - 3 H maps f to a meaning m'



The naming game

- **Goal:**
 - optimal communication
 - both want to minimize the distance between m and m'
- **Strategies:**
 - speaker: mapping S from M to F
 - hearer: mapping H from F to M
- **Average utility:** (identical for both players)

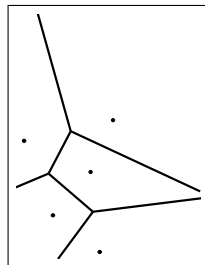
$$u(S, H) = \int_M p_m \times \exp(-d(m, H(S(m))))^2 dm$$

vulgo: average similarity between speaker's meaning and hearer's meaning



Voronoi tessellations

- suppose H is given and known to the speaker: which speaker strategy would be the best response to it?
 - every form f has a “prototypical” interpretation: $H(f)$
 - for every meaning m : S 's best choice is to choose the f that minimizes the distance between m and $H(f)$
 - optimal S thus induces a **partition** of the meaning space
 - Voronoi tessellation, induced by the range of H



Okabe et al. (1992) prove the following lemma (quoted from Gärdenfors 2000):

Lemma

The Voronoi tessellation based on a Euclidean metric always results in a partitioning of the space into convex regions.



ESSs of the naming game

- best response of H to a given speaker strategy S not as easy to characterize
- general formula

$$H(f) = \arg \max_m \int_{S^{-1}(f)} p_{m'} \times \exp(-d(m, m')^2) dm'$$

- such a hearer strategy always exists
- linguistic interpretation: H maps every form f to the **prototype** of the property $S^{-1}(f)$



ESSs of the naming game

Lemma

In every ESS $\langle S, H \rangle$ of the naming game, the partition that is induced by S^{-1} on M is the Voronoi tessellation induced by $H[F]$.



ESSs of the naming game

Lemma

In every ESS $\langle S, H \rangle$ of the naming game, the partition that is induced by S^{-1} on M is the Voronoi tessellation induced by $H[F]$.

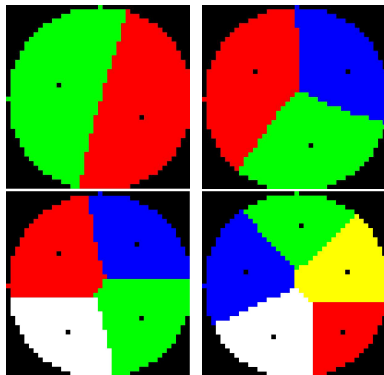
Theorem

For every form f , $S^{-1}(f)$ is a convex region of M .



Simulations

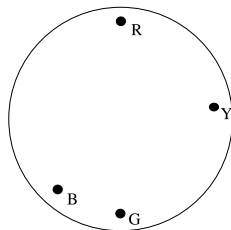
- two-dimensional circular meaning space
- discrete approximation
- uniform distribution over meanings
- initial strategies are randomized
- update rule according to (discrete time version of) replicator dynamics



A toy example

- suppose
 - circular two-dimensional meaning space
 - four meanings are highly frequent
 - all other meanings are negligibly rare
- let's call the frequent meanings
Red, Green, Blue and Yellow

$$p_i(\text{Red}) > p_i(\text{Green}) > p_i(\text{Blue}) > p_i(\text{Yellow})$$

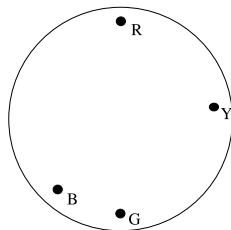


A toy example

- suppose
 - circular two-dimensional meaning space
 - four meanings are highly frequent
 - all other meanings are negligibly rare
- let's call the frequent meanings
Red, Green, Blue and Yellow

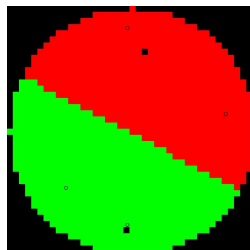
$$p_i(\text{Red}) > p_i(\text{Green}) > p_i(\text{Blue}) > p_i(\text{Yellow})$$

Yes, I made this up without empirical justification.



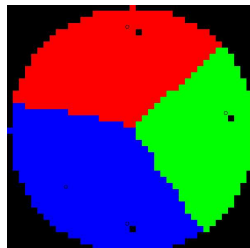
Two forms

- suppose there are just two forms
- only one Strict Nash equilibrium (up to permutation of the forms)
- induces the partition $\{\mathbf{Red}, \mathbf{Blue}\} / \{\mathbf{Yellow}, \mathbf{Green}\}$



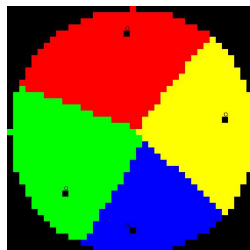
Three forms

- if there are three forms
- two Strict Nash equilibria (up to permutation of the forms)
- partitions $\{\text{Red}\}/\{\text{Yellow}\}/\{\text{Green, Blue}\}$ and $\{\text{Green}\}/\{\text{Blue}\}/\{\text{Red, Yellow}\}$
- only the former is **stochastically stable** (resistent against random noise)



Four forms

- if there are four forms
- one Strict Nash equilibrium (up to permutation of the forms)
- partitions
 $\{\text{Red}\}/\{\text{Yellow}\}/\{\text{Green}\}/\{\text{Blue}\}$



Meaning spaces

- assumption: utility is correlated with similarity between speaker's meaning and hearer's meaning
- consequences:
 - convexity of meanings
 - prototype effects
 - uneven probability distribution over meanings leads to the kind of implicational universals that are known from typology of color terms



EGT and language evolution

- EGT is well-suited to model utterance based, horizontal cultural language evolution
- allows to characterize attractor states in a static way, regardless of the micro-structure of language change
- possible refinements
 - stochastic evolution
 - spatial/network structure between agents



- Gärdenfors, P. (2000). *Conceptual Spaces*. The MIT Press, Cambridge, Mass.
- Okabe, A., B. Boots, and K. Sugihara (1992). *Spatial tessellations: concepts and applications of Voronoi diagrams*. Wiley, Chichester.

