Mathematical and computational models of language evolution

Gerhard Jäger

Institute of Linguistics, Tübingen University

DGfS Summer School

August 13, 2013

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Universal Darwinism

Necessary aspects of biological evolution

- population
- replication
- inheritance of differential features
- variation within the population with respect to heritable features
- differential fitness, i.e. success in replication
- fitness is correlated with heritable features
- therefore: natural selection, genetic drift

Contingent aspects of differential features

- dichotomy geno-type/phenotype
- discrete units of inheritance (genes/alleles)
- non-directedness of mutations
- non-heredity of acquired features
- mutations

Universal Darwinism

- any system exhibiting the necessary aspects can be considered to undergo evolution
- note that this is a definition, not an empirical fact
- Universal Darwinism has been applied (a.o.) to
  - ideas (*memes*)
  - religions
  - computer programs
  - companies
  - quantum states
  - universes
  - robots
  - neurons
  - technologies
  - ...

Cultural language evolution

There are (at least) three ways how languages/linguistic entities can be considered an evolutionary system:

1. evolution of the human language faculty (cf. course by Tallerman)
   - ordinary biological evolution of human populations under the aspect of differential fitness being determined by innate propensity for language learning/usage
   - individuals: humans
   - replication: biological reproduction
2. co-evolution of languages and genes
   - individuals: I-languages (each one living in the brain of a single human)
   - replication: language acquisition
   - fitness is strongly tied to biological fitness
   - explains co-evolution of linguistic and genetic characteristics (cf. course by Dediu & Vernes)

Cultural language evolution

3. language change, seen as an evolutionary process (this course)
   - individuals: **linguemes** = pieces of linguistic structure that are memorized/imitated as a unit
   - population: E-language
   - replication: imitation (with first language acquisition as a special case)
   - explains
     - universal features of natural languages that are not genetically determined
     - patterns of language change

Language evolution

"The formation of different languages and of distinct species, and the proofs that both have been developed through a gradual process, are curiously parallel. . . . Max Müller has well remarked: 'A struggle for life is constantly going on amongst the words and grammatical forms in each language. The better, the shorter, the easier forms are constantly gaining the upper hand, and they owe their success to their inherent virtue.' To these important causes of the survival of certain words, mere novelty and fashion may be added; for there is in the mind of man a strong love for slight changes in all things. The survival or preservation of certain favoured words in the struggle for existence is natural selection." (Darwin 1871:465f.)

Language evolution

standard assumptions about prerequisites for evolutionary processes (see for instance Richard Dawkins' work)

- population of **replicators** (for instance genes)
- (almost) faithful replication (for instance DNA copying)
- variation
- differential replication $\rightsquigarrow$ selection

Language evolution

modes of linguistic replication

- the biological inheritance of the human language faculty,
- first language acquisition, which amounts to a vertical replication of language competence from parents (or, more generally, teachers) to infants, and
- imitation of certain aspects of language performance in language usage (like the repetition of words and constructions, imitation of phonetic idiosyncrasies, priming effects etc.)

Language evolution

What are the replicators?

- I-languages/grammars?
- E-languages/grammars?
- linguemes?
- rules?
- utterances (or features thereof)?

Perhaps Dawkins' conceptual framework is too narrow...

George R. Price

- 1922–1975
- studied chemistry; briefly involved in Manhattan project; lecturer at Harvard
- during the fifties: application of game theory to strategic planning of U.S. policy against communism
  - proposal to buy each Soviet citizen two pair of shoes in exchange for the liberation of Hungary
- tried to write a book about the proper strategy to fight the cold war, but *"the world kept changing faster than I could write about it"*, so he gave up the project
- 1961–1967: IBM consultant on graphic data processing

George R. Price

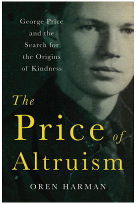
- 1967: emigration to London (with insurance money he received for medical mistreatment that left his shoulder paralyzed)
- 1967/1968: freelance biomathematician

George R. Price

- discovery of the **Price equation**
- leads to an immediate elegant proof of **Fisher's fundamental theorem**
- invention of **Evolutionary Game Theory**
    - Manuscript *Antlers, Intraspecific Combat, and Altruism* submitted to *Nature* in 1968; contained the idea of a mixed ESS in the Hawk-and-Dove game
    - accepted under the condition that it is shortened
    - reviewer: John Maynard Smith
    - Price never resubmitted the manuscript, and he asked Maynard Smith not to cite it
    - 1972: Maynard Smith and Price: *The Logic of Animal Conflict*
    - Price to Maynard Smith: *"I think this the happiest and best outcome of refereeing I've ever had: to become co-author with the referee of a much better paper than I could have written by myself."*

George R. Price

- 1968–1974: honorary appointment at the Galton Labs in London
- 1970: conversion to Christianity; after that, most of his attention was devoted to biblical scholarship and charity work
- around 1971: *The Nature of Selection* (published posthumously in 1995 in Journal of Theoretical Biology)
- early 1975: suicide



The Nature of Selection

"A model that unifies all types of selection (chemical, sociological, genetical, and every other kind of selection) may open the way to develop a general 'Mathematical Theory of Selection' analogous to communication theory."

The Nature of Selection

"Selection has been studied mainly in genetics, but of course there is much more to selection than just genetical selection. In psychology, for example, trial-and-error learning is simply learning by selection. In chemistry, selection operates in a recrystallisation under equilibrium conditions, with impure and irregular crystals dissolving and pure, well-formed crystals growing. In palaeontology and archaeology, selection especially favours stones, pottery, and teeth, and greatly increases the frequency of mandibles among the bones of the hominid skeleton. In linguistics, selection unceasingly shapes and reshapes phonetics, grammar, and vocabulary. In history we see political selection in the rise of Macedonia, Rome, and Muscovy. Similarly, economic selection in private enterprise systems causes the rise and fall of firms and products. And science itself is shaped in part by selection, with experimental tests and other criteria selecting among rival hypotheses."

The Nature of Selection

Concepts of selection

- subset selection
- Darwinian selection

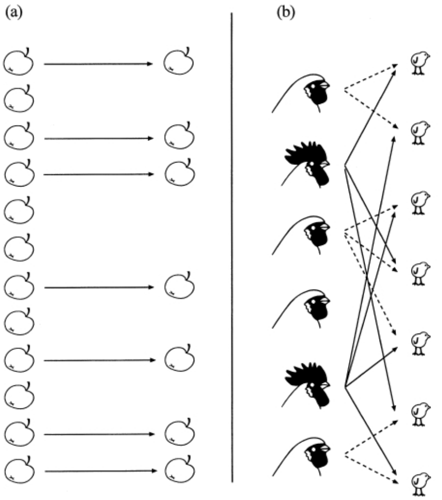


FIG. 1. Conventional concepts of selection. (a) Subset selection. (b) Darwinian selection.

The Nature of Selection

Concepts of selection

- common theme:
  - two time points
    - t: population before selection
    - t': population after selection

- partition of populations into N bins
- parameters
  - abundance w_i/w_i' of bin i before/after selection
  - quantitative character x_i/x_i' of each bin

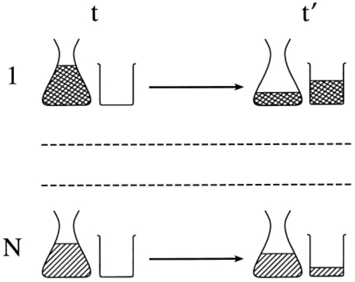


FIG. 2. A solution selection example.

The Nature of Selection

- each individual at t' **corresponds** to exactly one item at t
- nature of correspondence relation is up to the modeler — biological descendance is an obvious, but not the only possible choice
- partition of t-population induces partition of t'-population via correspondence relation

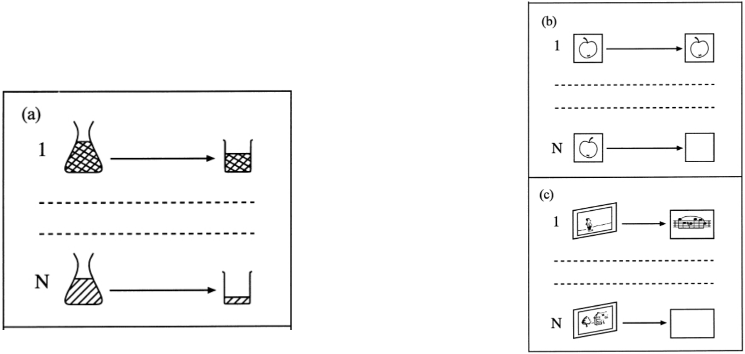
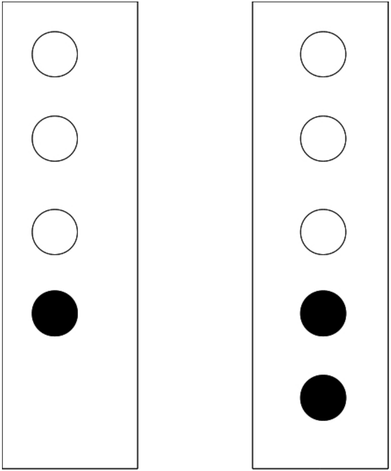


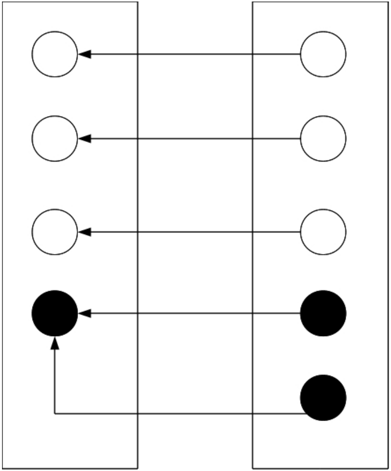
FIG. 3. Three selection examples arranged in the pattern of the general selection model. (a) The essential elements of the Fig. 2 example. (b) How the Fig. 1(a) example is fitted to the general model.

Schematic example



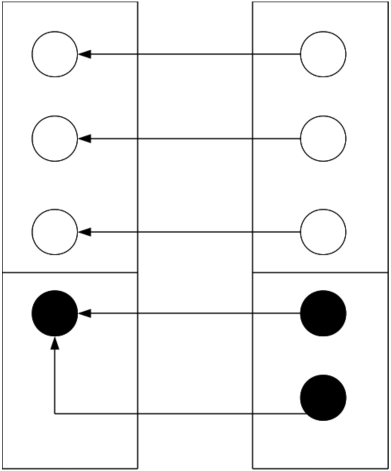
population at two points in time

Schematic example



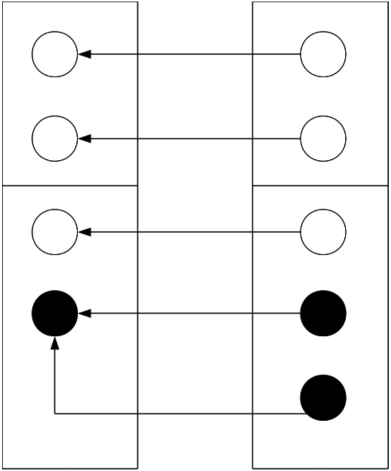
adding correspondence relation

Schematic example



adding partition structure

Schematic example



adding partition structure

The Nature of Selection

property change

- quantitative character x may be different between parent and offspring
- $\Delta x_i = x'_i - x_i$ need not equal 0
- models unfaithful replication (e.g. mutations in biology)

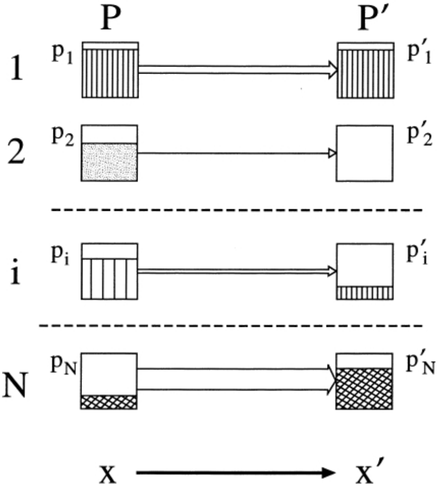
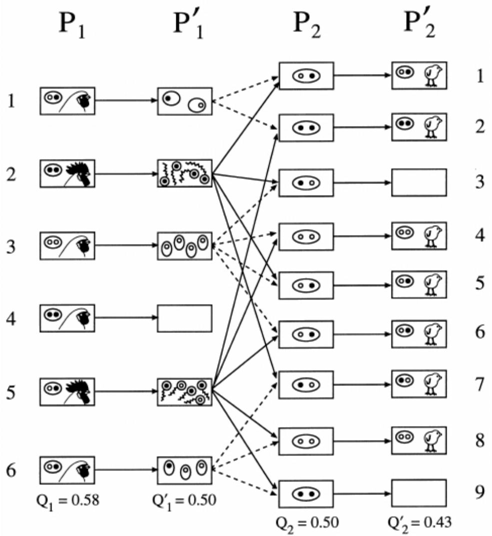


FIG. 4. The general selection model.

The Nature of Selection

genetical selection:



The Price equation

Parameters

- w_i: abundance of bin i in old population
- w_i': abundance of descendants of bin i in new population
- $f_i = w_i'/w_i$: fitness of type-i individuals
- $f = \frac{\sum_i w_i'}{\sum_i w_i}$: fitness of entire population
- x_i: average value of x within i-bin
- x_i': average value of x within descendants of i-bin
- $\Delta x_i = x_i' - x_i$: change of x_i
- $x = \sum_i \frac{w_i}{w} x_i$: average value of x in old population
- $x' = \sum_i \frac{w_i'}{w} x_i'$: average value of x in new population
- $\Delta x = x' - x$: change of expected value of x

The Price equation

Discrete time version

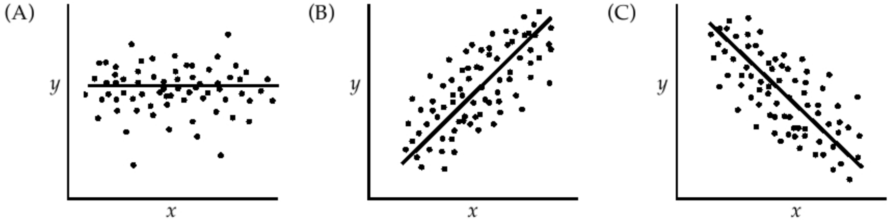
$$f\Delta x = Cov(f_i, x_i) + E(f_i \Delta x_i)$$

- $Cov(f_i, x_i)$: change of x due to natural selection
- $E(f_i \Delta x_i)$: change of x due to unfaithful replication

Continuous time version

$$\dot{E}(x) = Cov(f_i, x_i) + E(\dot{x}_i)$$

The Price equation

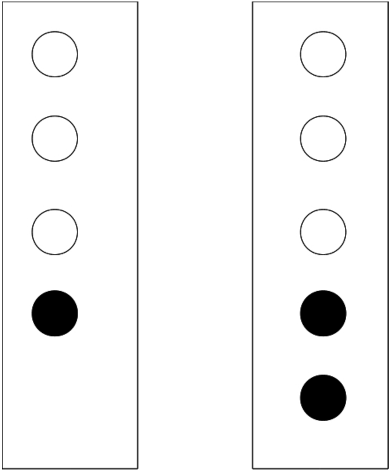


- Covariance $\approx$ slope of linear approximation
  - (A) $= 0$: no dependency between x and y
  - (B) > 0: high values of x correspond, on average, to high values of y and vice versa
  - (C) < 0: high values of x correspond, on average, to low values of y and vice versa

The Price equation

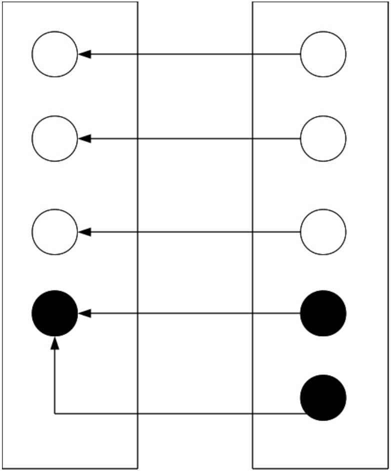
- important: the equation is a tautology
- follows directly from the definitions of the parameters involved
- very general; no specific assumptions about the nature of the replication relation, the partition of population into bins, the choice of the quantitative parameter under investigation
- many applications, for instance in investigation of group selection

Schematic example



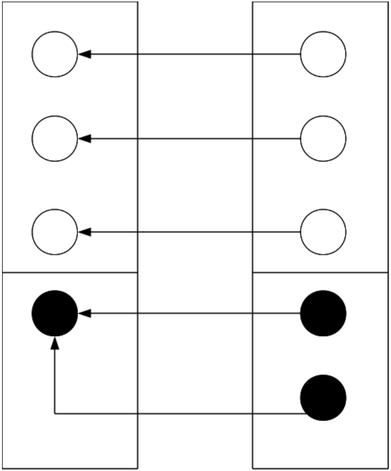
population at two points in time

Schematic example



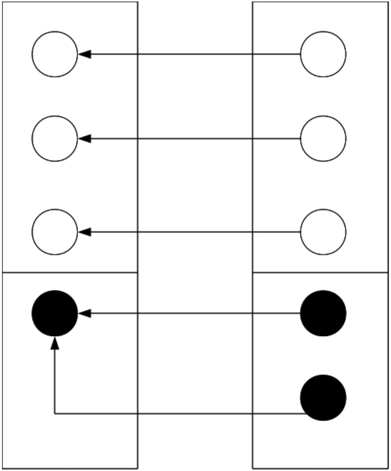
adding correspondence relation

Schematic example



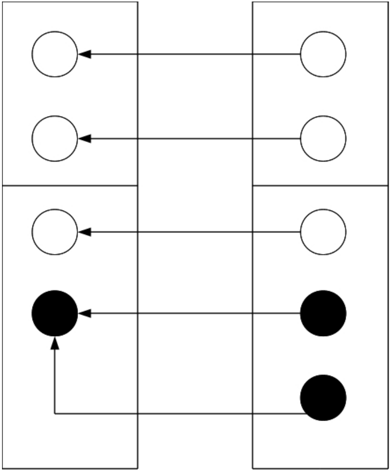
adding partition structure

Schematic example



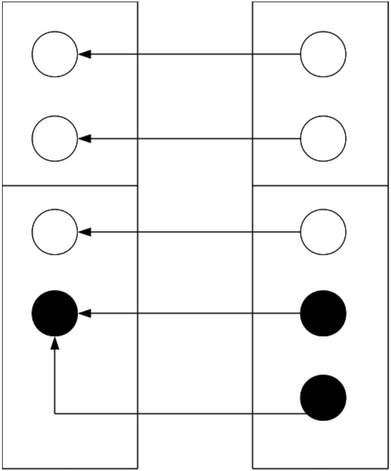
$$\begin{aligned} f\Delta x &= {\color{red}Cov(f_i, x_i)} + {\color{blue}E(f_i\Delta x_i)} \\ 0.1875 &= {\color{red}0.1875} + {\color{blue}0} \end{aligned}$$

Schematic example



adding a different partition structure

Schematic example



$$
\begin{aligned}
f\Delta x &= {\color{red}Cov(f_i, x_i)} + {\color{blue}E(f_i\Delta x_i)}\\
0.1875 &= {\color{red}0.0625} + {\color{blue}0.125}
\end{aligned}
$$

Applications of the Price equation

Fisher's Theorem

- x can be any quantitative character, including fitness
- for $x = f$, we have

$$\dot{f} = Var_i(f_i) + E_i(\dot{f}_i)$$

  - $Var_i(f_i)$: increase in average fitness due to natural selection
  - $E_i(\dot{f}_i)$: decrease in average fitness due to deterioration of the environment

Applications of the Price equation

$$\dot{E}(x) = Cov(f_i, x_i) + E(\dot{x}_i)$$

Group selection

- population of groups that each consists of individuals
- bins = groups
- first term:
  - covariance between a certain trait x and group fitness
  - corresponds to natural selection at the group level
- second term:
  - avarage change of x **within** group
  - corresponds to natural selection at the individual level
- for "altruistic" traits, first term would be positive but second term negative

Consequences of Price's approach

- no single "correct" way to model language evolution
- prerequisites for applying Price's approach:
  - two populations at different time points
  - natural assignment of items of the new population to items in the old population
- it is up to the model builder
  - what populations consist of (any measurable set would do)
  - the evolution of which character is studied (as long as it is quantitative in nature)
  - what the nature of the "replication" relation is — any function from the new population to the old one will do
  - how populations are partitioned into bins

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

Exemplar dynamics and blending inheritance

Model architecture (inspired by Wedel)

- exemplars are n-dimensional vectors ($n = 2$ in the sample simulation)
- exemplar pool is initialized with random set
- creation of new exemplars:
  - draw a sample S of s exemplars at random from the exemplar pool
  - find the mean m of S

$$m = \frac{1}{s} \sum_{v \in S} v$$

  - add m to exemplar pool and forget oldest exemplar

Exemplar dynamics and blending inheritance

Assumptions

- population of exemplars is practically infinite
- continuous distribution over some finite vector space
- all exemplars are equally likely to be picked out as part of S

[simulation]

Modeling decisions

- ancestor population: old exemplar pool
- successor population: new exemplar pool, including the newly created exemplar
- all elements of S are “parents” of the newly added exemplar
- each exemplar forms its own bin

Exemplar dynamics and blending inheritance

Consequences

- all bins have identical fitness
- first term of the Price equation can be ignored
- continuous population $\rightarrow$ continuous time dynamics

$$\dot{E}(x) \quad = \quad E(\dot{x}_i)$$

Exemplar dynamics and blending inheritance

First application: evolution of the population average

- let g be the center of gravitation of the population
- character to be studied: v_i, i.e. position of the i-the exemplar
- then

$$\dot{v}_i = (1-\alpha)v_i + \alpha(g - v_i)$$

- hence:

$$\dot{E}(v_i) = \dot{g} = 0$$

- in words: the center of gravitation remains constant

Exemplar dynamics and blending inheritance

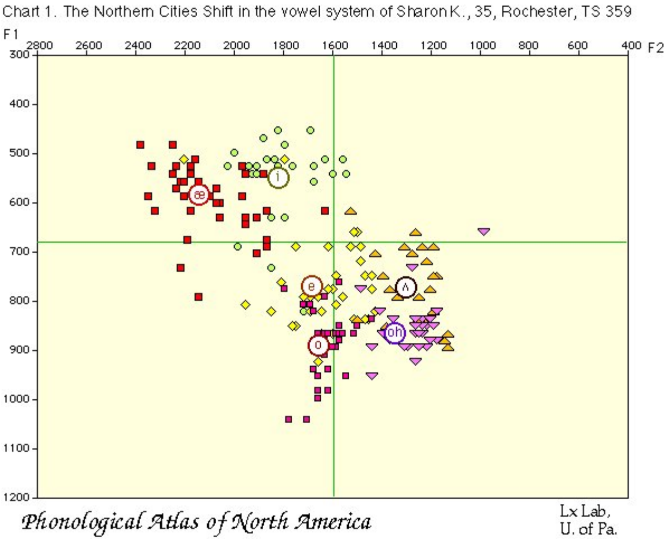
Second application: evolution of variance

- character to be studied: variance of the population

$$
\begin{aligned}
Var(v_i) &= E[(v_i - g)^2] \\
\dot{Var}(v_i) &= E[(v_i \dot{-} g)^2] \\
\dot{Var}(v_i) &= -\beta Var(v_i) \\
Var(v_i)(t) &= k \exp(-\beta t)
\end{aligned}
$$

- in words: the variance decreases at exponential rate

Noisy exemplar communication



Noisy exemplar communication

Setup

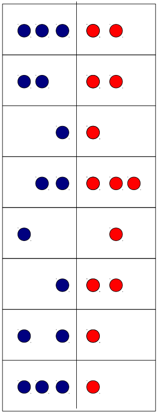
- pool of exemplars A
- each exemplar has category labels (for instance phonemes)
- each exemplar has coordinates in some signal space (for instance F1/F2 for vowels)
- confusion matrix Z between signals

Noisy exemplar communication

Dynamics

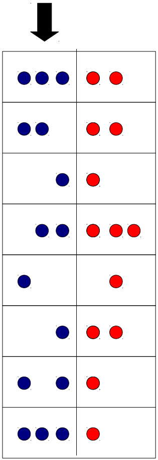
- nature picks out a category c at random
- player picks at random an exemplar with category label c and sends it; let v be the signal component of this exemplar
- a possibly different signal v' is received (with probability $Z_{vv'}$)
- among the exemplars with signal component v', one is picked out at random
- if its category component $= c$, (c, v) and (c, v') are added to the exemplar pool

Noisy exemplar communication



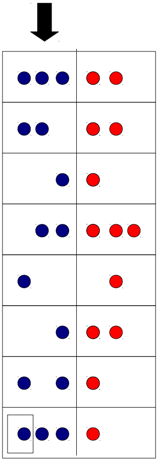
exemplar pool

Noisy exemplar communication



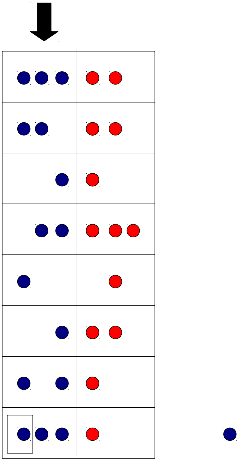
nature picks a category

Noisy exemplar communication



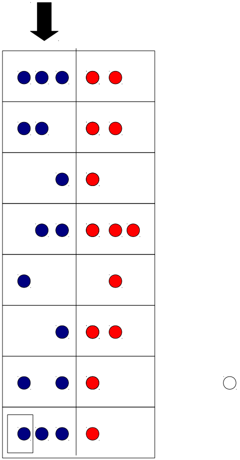
agent picks an exemplar of that category

Noisy exemplar communication



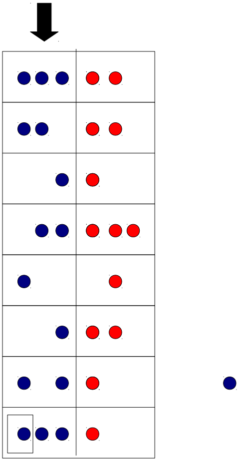
agent emits signal corresponding to that exemplar

Noisy exemplar communication



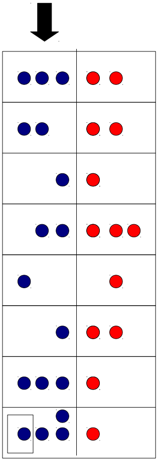
agent receives a (possibly different) signal

Noisy exemplar communication



agent guesses a category based on exemplar pool

Noisy exemplar communication



if successful, both sent and received exemplar are memorized

Price style modeling

$$
\begin{aligned}
\dot{E}(\delta(c,v)) &= Cov(\delta, f) + E(\dot{\delta}) \\
\dot{a}_{c^*v^*} &= 2a_{c^*v^*}(u(c^*,v^*) - \tilde{u}) + \sum_v a_{c^*v} u(c^*,v) z_{vv^*} - a_{c^*v^*} u(c^*,v^*) \\
u(c,v) &= \frac{p^*(c)}{\sum_{v'} a_{csv'} (ZR)_{vc}} \\
\tilde{u} &= \sum_{c,v} a_{cv} u(c,v) \\
r_{vc} &= \frac{a_{cv}}{\sum_{v'} a_{v'c}}
\end{aligned}
$$

[simulation]