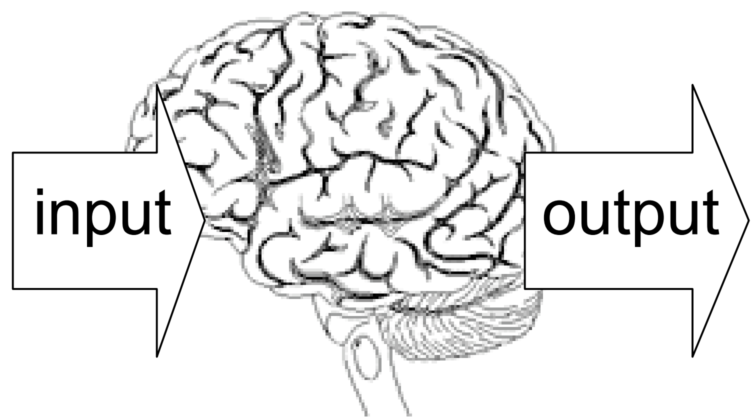
Human and machine learning

Tom Griffiths
Department of Psychology
Cognitive Science Program
University of California, Berkeley

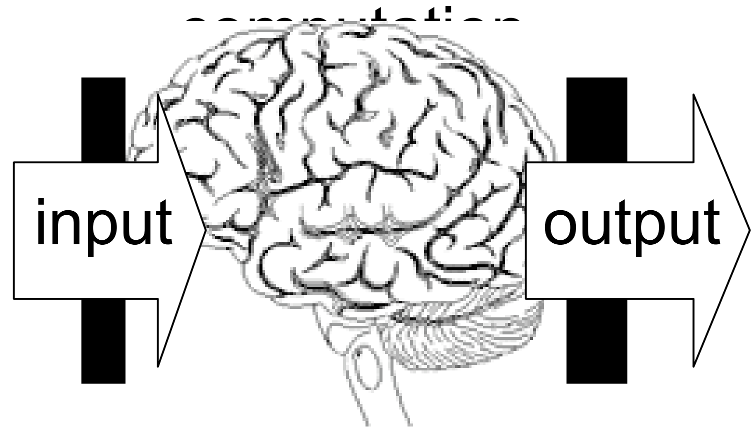
Computation

Cognition

Information processing



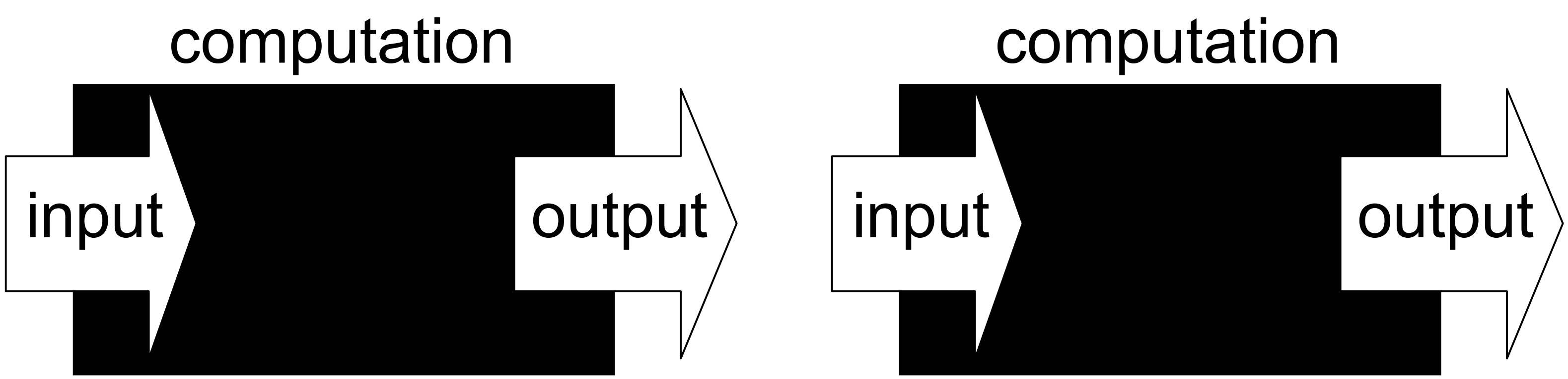
Information processing



Convergent evolution

Computers and brains face similar problems…

Do they use similar solutions?



Outline

Spam filters and classification

Search and memory

Inferring preferences

Outline

Spam filters and classification

Search and memory

Inferring preferences

Spam

= Cyrillic?

= has “!”?

= links?

= CAPS?

From: Надя <haxilyki@whayne.com>
Subject: **[SPAM:###] Мириады чувственности**
Date: February 26, 2009 8:24:43 AM PST
To: Tom Griffiths
Reply-To: haxilyki@whayne.com

МЕ284 Вот это да!
ПЧ607 Такие необыкновенные женщины
ВК468 Они невероятно чувственные
ВД163 Они способны разбудить желание в любом мужчине
ГХ752 Хочешь проверить?

From: Renning Fieldson <peroxisomal@austexdies.com>
Subject: **[SPAM:####] [SPAM:#####] More orgasmss**
Date: February 25, 2009 3:34:45 PM PST
To: Tom Griffiths
Reply-To: Renning Fieldson <peroxisomal@austexdies.com>

New OOrgasm Enhancer
Click HERE

Him from his mind, went to work on his favourite after the vote result is posted to news.announce.newgroups, clothes on a neglected bed, and its pillow was he instantly saw that it would be impossible for it out of his hide.' illustration: lincoln and.

From: Staci Malone <gershon@psych.stanford.edu>
Subject: **[SPAM:####] [SPAM:######] Show your friends how filthy rich You are**
Date: February 23, 2009 9:56:40 PM PST
To: Jillian Irwin <gershon@psych.stanford.edu>

Loving yourself is the first step in loving life. And what better way to do it, than by getting
http://nocefawef.cn

Now that the Holidays are behind us and stores everywhere are offering their lowest pric
distinguished watch at a ridiculously low price!
http://nocefawef.cn

Don't delay your pleasure: our incredible watch collection awaits you at Exquisite Reps,

Not spam

From: National Science Foundation Update <nsf-update@nsf.gov>
Subject: **How to Teleport Quantum Information from One Atom to Another**
Date: February 26, 2009 5:40:02 AM PST
To: Tom Griffiths
Reply-To: National Science Foundation Update <nsf-update@nsf.gov>

How to Teleport Quantum Information from One Atom to Another

Researchers have shown for the first time how to use a process ca
More at http://www.nsf.gov/discoveries/disc_summ.jsp?cntn_id=1

This is an NSF Discoveries item.

This e-mail update was generated automatically based on your subscription to the
messages.

You can adjust your National Science Foundation Update subscriptions or delivery
stop subscriptions on this page. If you have questions or problems with National Sc

National Science Foundation · 4201 Wilson Boulevard · Arlington, VA 22230 · 703-292-5111

From: ABC NewsMail <newslists@your.abc.net.au>
Subject: **ABC NewsMail - morning edition - text only**
Date: February 25, 2009 1:10:00 PM PST
To: Tom Griffiths

ABC News
Thursday February 26, 2009
(For more news visit ABC News Online at http://abc.net.au/news/)

To receive this email in HTML with your preferred topics, log in with your email address at:
http://abc.net.au/news/alerts/default.htm

ABC NewsMail headlines at a glance

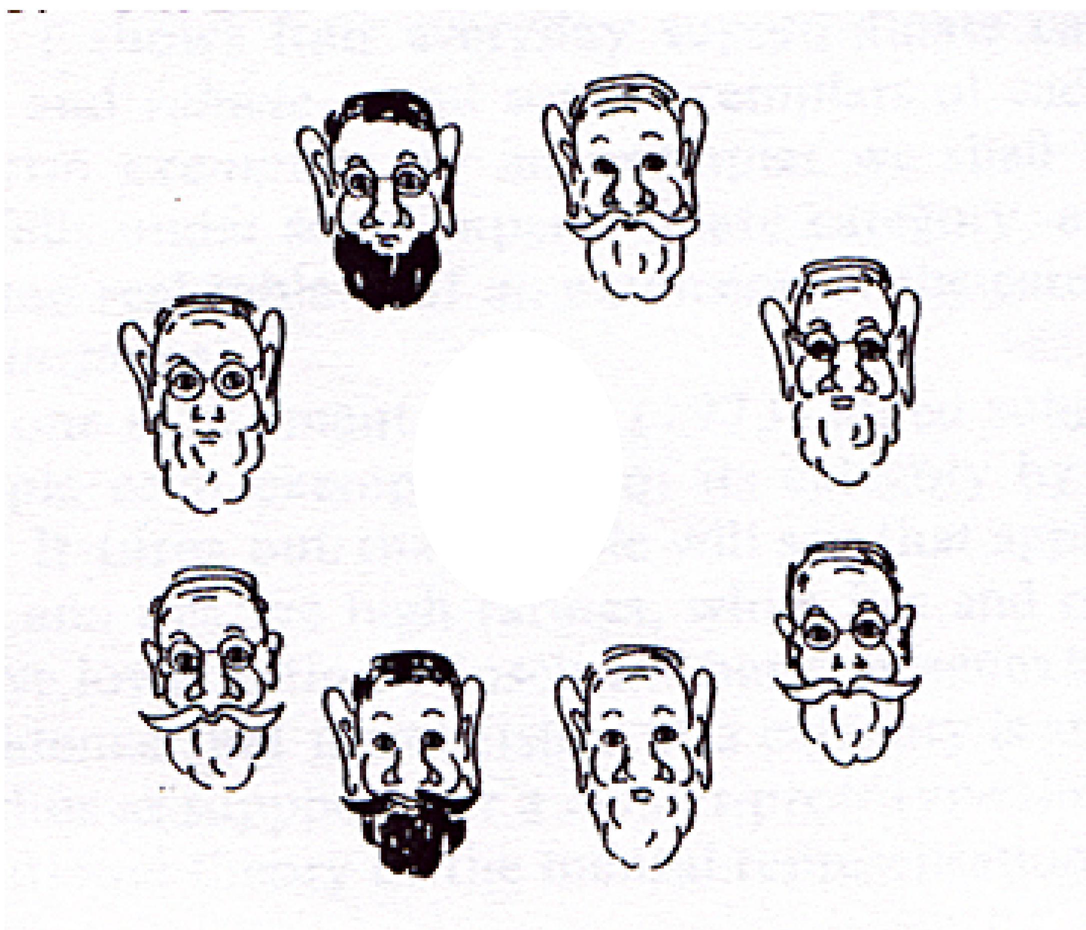
9 dead, more escape, after Amsterdam jet crash

Coroner to rule on Beaconsfield collapse

Mumbai siege gunman charged with 'waging war'

Hot conditions to test crews on 1,000k fire front

Family resemblance



Bayes' rule

Posterior probability

Likelihood

Prior probability

$$P(h \mid d) = \frac{P(d \mid h) P(h)}{\sum_{h' \in H} P(d \mid h') P(h')}$$

Sum over space of hypotheses

h: hypothesis
d: data

Bayesian inference

$$P(c \mid x) = \frac{P(x \mid c)P(c)}{\sum_c P(x \mid c)P(c)}$$ simplifies for 2 categories

$$\frac{P(A \mid x)}{P(B \mid x)} = \frac{P(x \mid A)}{P(x \mid B)} \frac{P(A)}{P(B)}$$ odds form

$$\log\frac{P(A \mid x)}{P(B \mid x)} = \log\frac{P(x \mid A)}{P(x \mid B)} + \log\frac{P(A)}{P(B)}$$ log odds form

$$\log\frac{P(A \mid x)}{P(B \mid x)} = \sum_{k=1}^{m} \log\frac{P(x_k \mid A)}{P(x_k \mid B)} + \log\frac{P(A)}{P(B)}$$ Naïve Bayes

A simple classifier

$$\log\frac{P(A \mid x)}{P(B \mid x)} = \sum_{k=1}^{m} \log\frac{P(x_k \mid A)}{P(x_k \mid B)} + \log\frac{P(A)}{P(B)}$$

x_k	$P(x_k \mid \text{spam})$	$P(x_k \mid \text{not spam})$
= Cyrillic?	high	low
= has "!"?	high	medium
= links?	high	medium
= CAPS?	medium	low
= has "Viagra"?	medium	low
= has "Science"?	low	medium

Coevolution

From: Renning Fieldson <peroxisomal@austexdies.com>
Subject: **[SPAM:####] [SPAM:#####] More orgasmss**
Date: February 25, 2009 3:34:45 PM PST
To: Tom Griffiths
Reply-To: Renning Fieldson <peroxisomal@austexdies.com>

New OOrgasm Enhancer
Click HERE

Him from his mind, went to work on his favourite after the vote result is posted to news.announce.newgroups, clothes on a neglected bed, and its pillow was he instantly saw that it would be impossible for it out of his hide.' illustration: lincoln and.

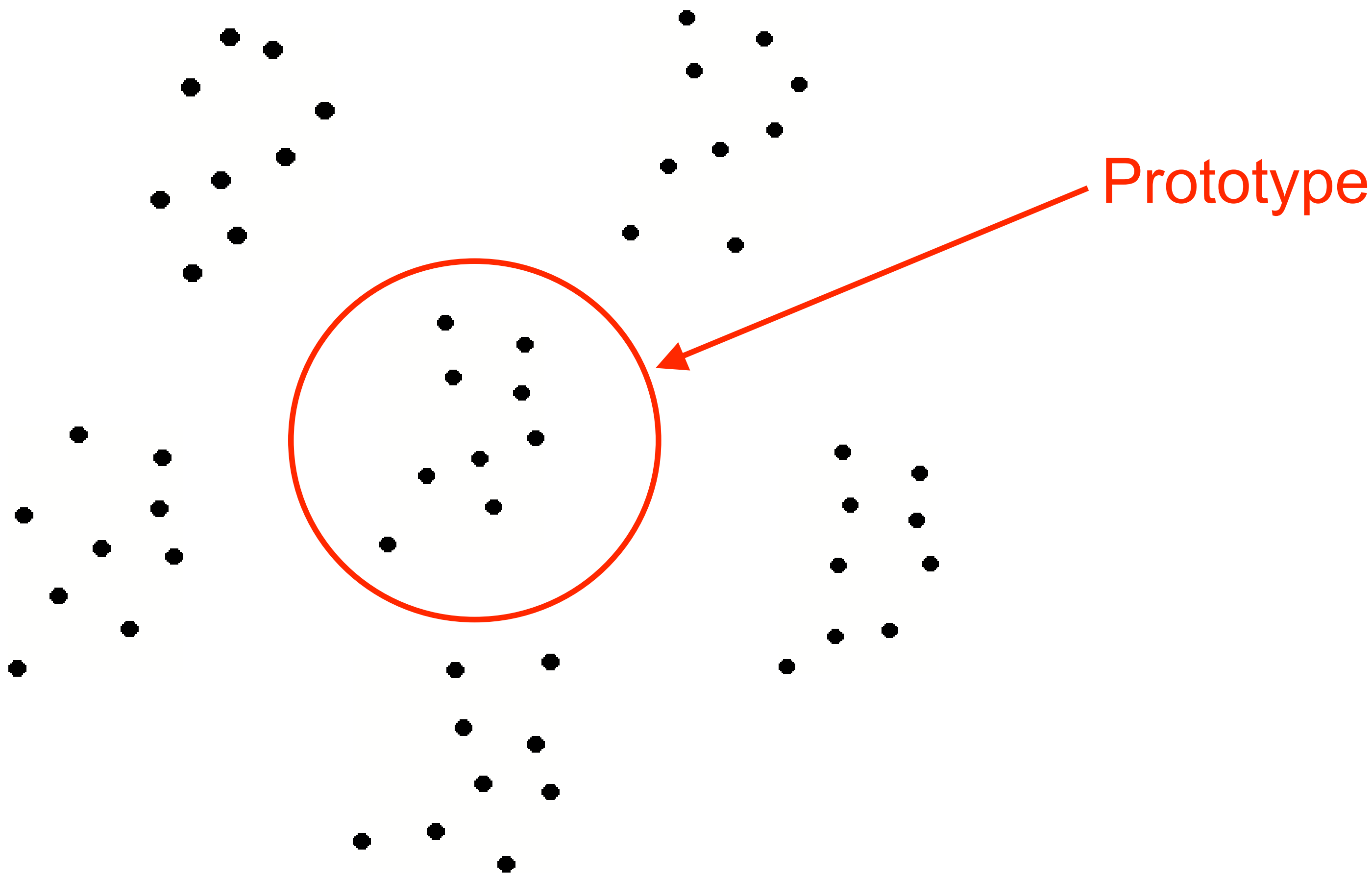
remove spam features

add non-spam features

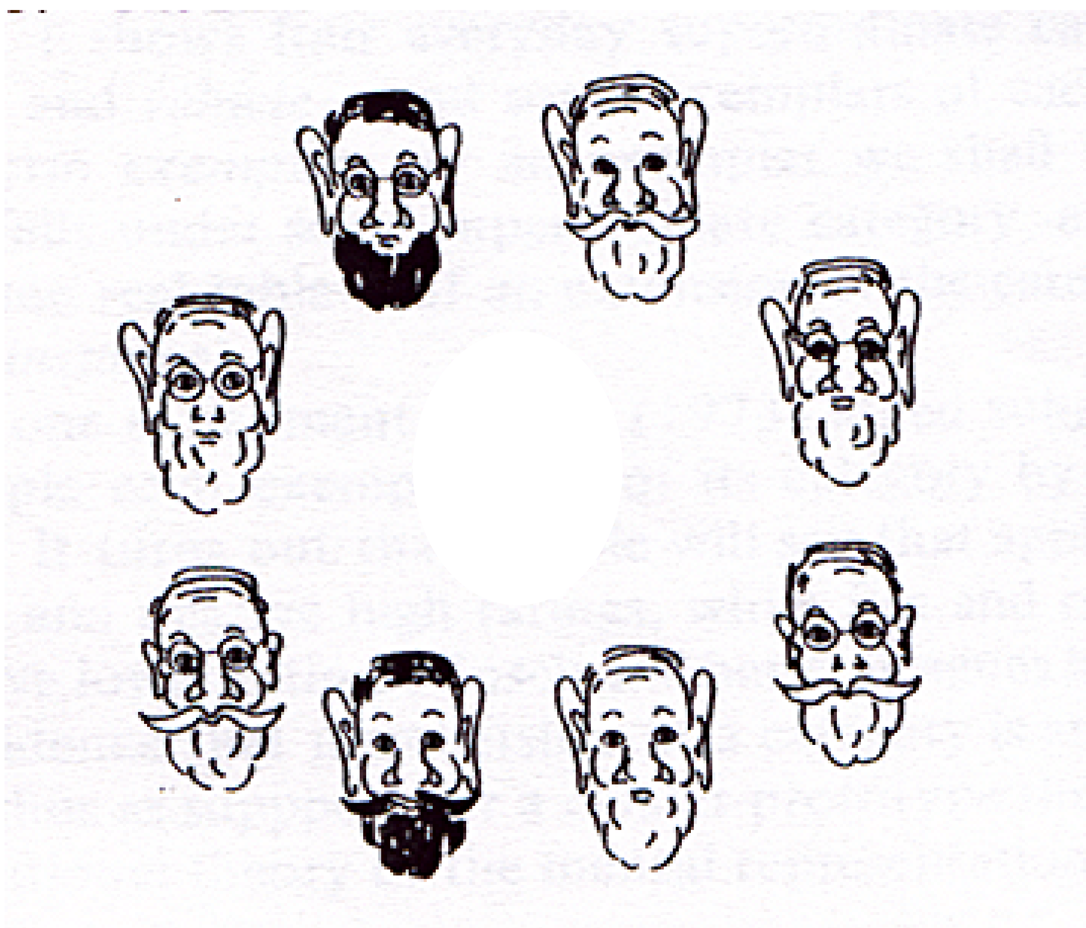
Categorization

cat $\Leftrightarrow$ small $\wedge$ furry $\wedge$ domestic $\wedge$ carnivore

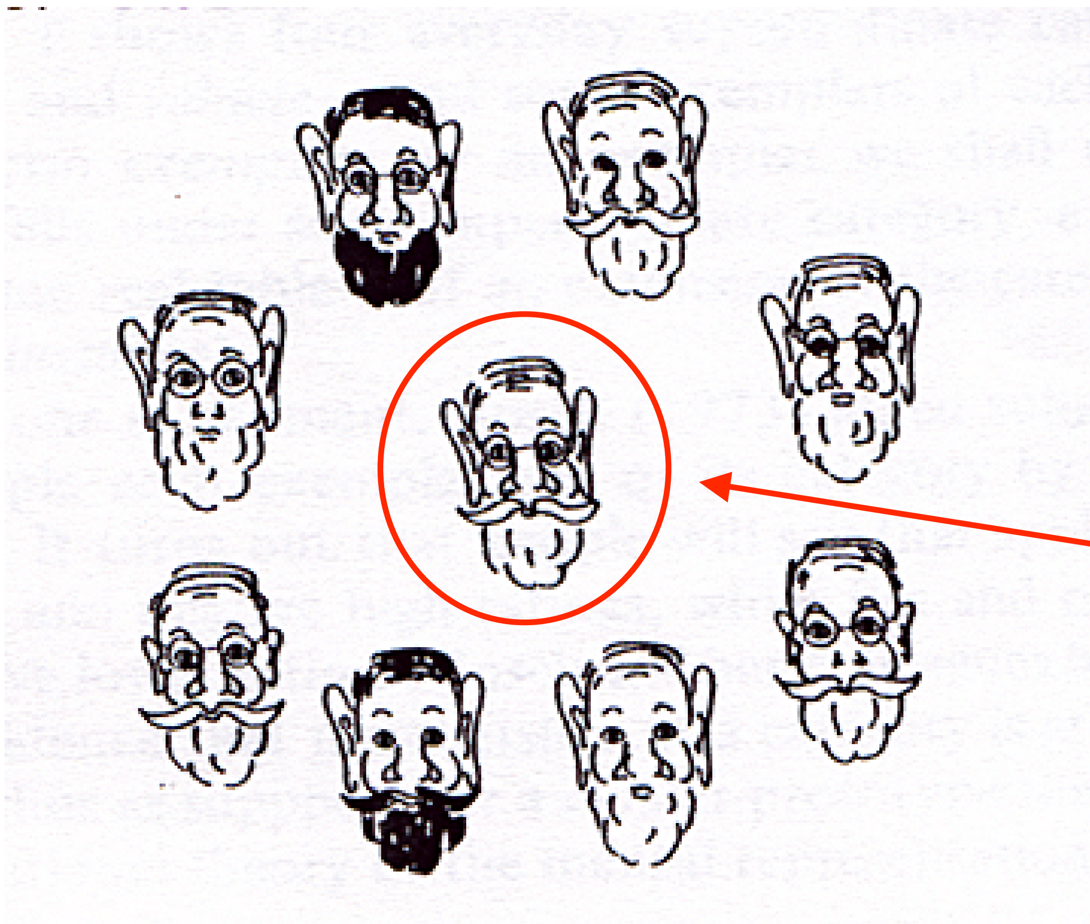
Posner & Keele (1968)



Family resemblance



Family resemblance



Prototype

Prototypes with features….

Prototype
e.g., binary vector with most frequent feature values

Distance
e.g., Hamming distance

$$d(x,\mu_A) = \sum_k |x_k - \mu_{A,k}|$$

choose category A if

$$d(x,\mu_A) < d(x,\mu_B)$$

$$\sum_k |x_k - \mu_{A,k}| < \sum_k |x_k - \mu_{B,k}|$$

Bayes and prototypes

$$\log \frac{P(A \mid x)}{P(B \mid x)} = \sum_{k=1}^{m} \log \frac{P(x_k \mid A)}{P(x_k \mid B)} + \log \frac{P(A)}{P(B)}$$

Naïve Bayes

choose category A if

$$P(A \mid x) > P(B \mid x)$$

$$\log \frac{P(A \mid x)}{P(B \mid x)} > 0$$

$$\sum_k \log \frac{P(x_k \mid A)}{P(x_k \mid B)} > 0$$

assuming $P(A) = P(B)$

$$\sum_k \log P(x_k \mid A) > \sum_k \log P(x_k \mid B)$$

define

$$P(x_k \mid A) = \begin{cases} 1-\varepsilon & x_k = \mu_{A,k} \\ \varepsilon & \text{otherwise} \end{cases}$$

$P(A|x) > P(B|x)$ if and only if…

$$\sum_k \left| x_k - \mu_{A,k} \right| < \sum_k \left| x_k - \mu_{B,k} \right|$$

Spam filters and classification

- A statistical analysis of the problem of classification yields a simple solution
  - weighted combination of features, with a threshold for final classification
- This solution is consistent with a theory of human category learning: prototypes
- Current research uses more sophisticated strategies to solve this problem, which also have analogues in human cognition

Outline

Spam filters and classification

Search and memory

Inferring preferences

Retrieving facts

Google

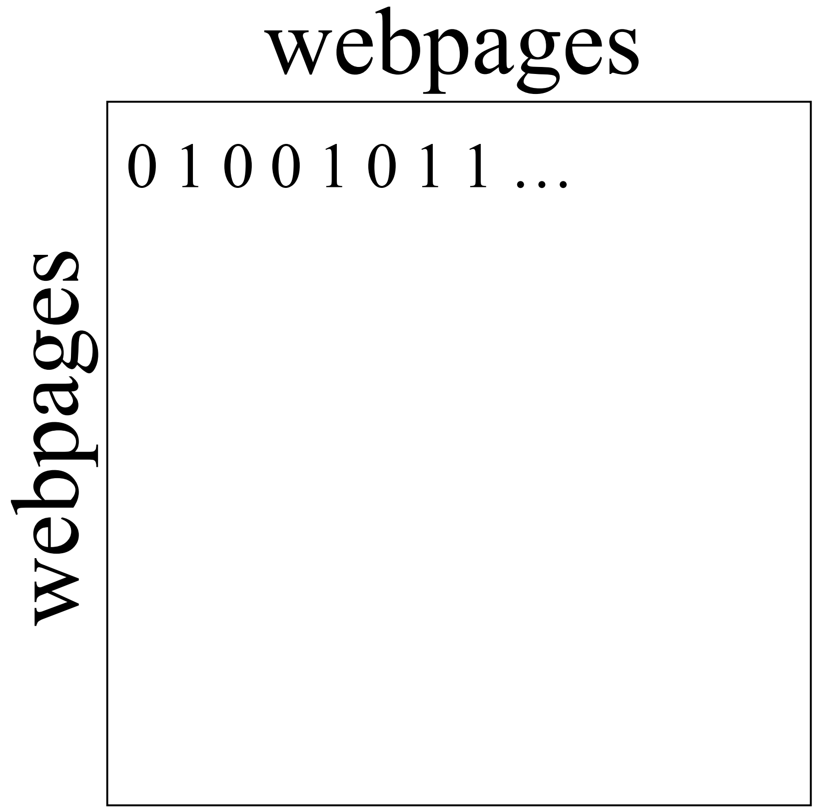
Bayes for search

- Data d are the terms of the query
- Hypotheses h are candidate webpages

- Assume likelihood $P(d|h)$ is constant for all webpages containing query, and 0 otherwise
  - posterior probabilities of matching webpages depend only on the prior…

- What prior $P(h)$ should we use?

Using link information



How Google works

(in 1998)

- Use $\mathbf{p}$ to denote the vector of importances of each of n webpages (one entry per webpage)

- Use $\mathbf{L}$ to denote the "link matrix", where $L_{ij} = 1$ if a link exists from j to i and 0 otherwise

- How should we define importance?
  - one option: number of pages linking to a page

$$p_j = \sum_i L_{ij} \qquad\qquad \mathbf{p} = \mathbf{L1}$$

How Google works
(in 1998)

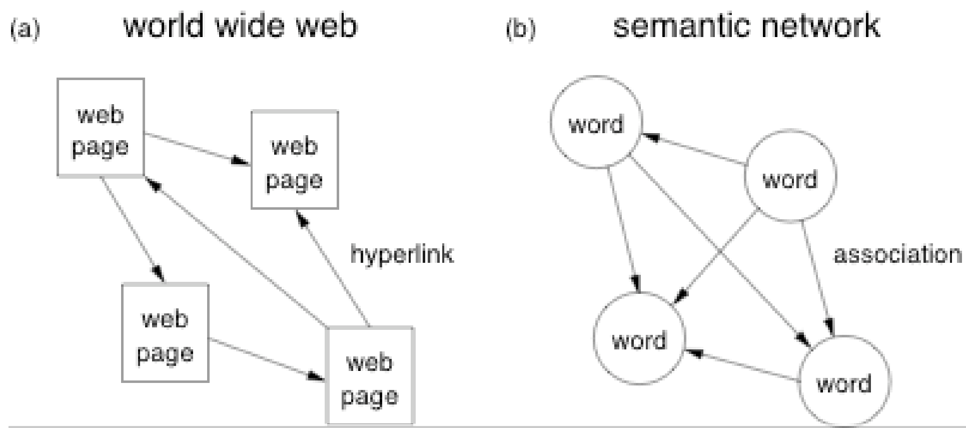
- A link from an important page should be worth more than a link from a less important page
- A link from a page making few links should be worth more than one making lots of links

- “PageRank” algorithm defines importance as

$$\mathbf{p} = \mathbf{M}\mathbf{p} \quad \text{where} \quad M_{ij} = \frac{1}{\sum_k L_{kj}}$$

An analogy…

- One model of knowledge: a semantic network



- Similar statistical properties to the web
  – short paths, power-law distributions, clustering
- Can we connect search to memory?

Word association

Cue:
PLANET

(Nelson, McEvoy & Schreiber, 1998)

Word association

Cue:

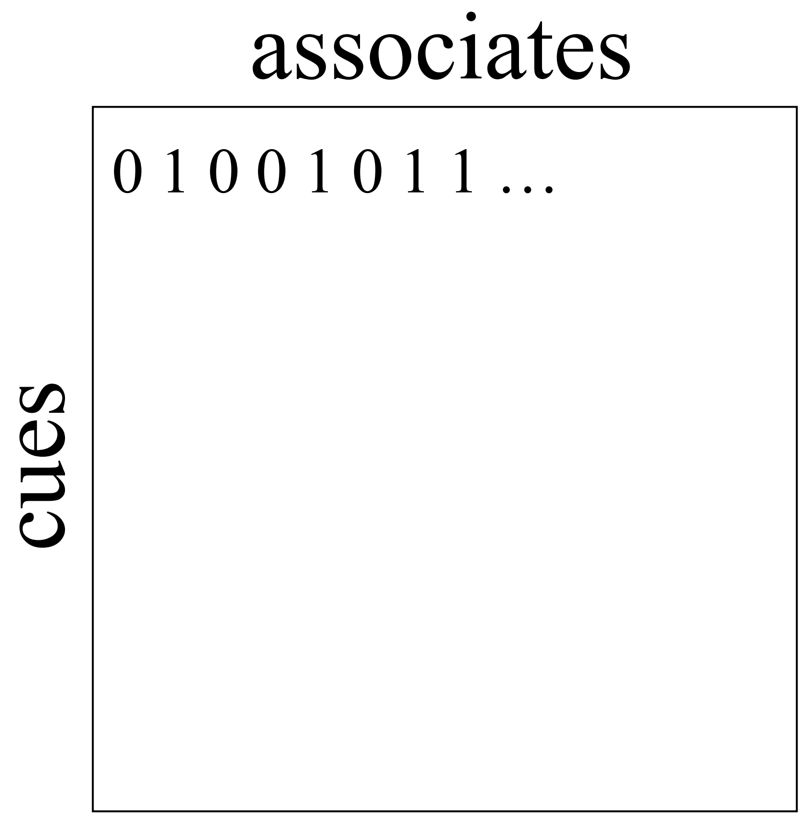
PLANET

Associates:

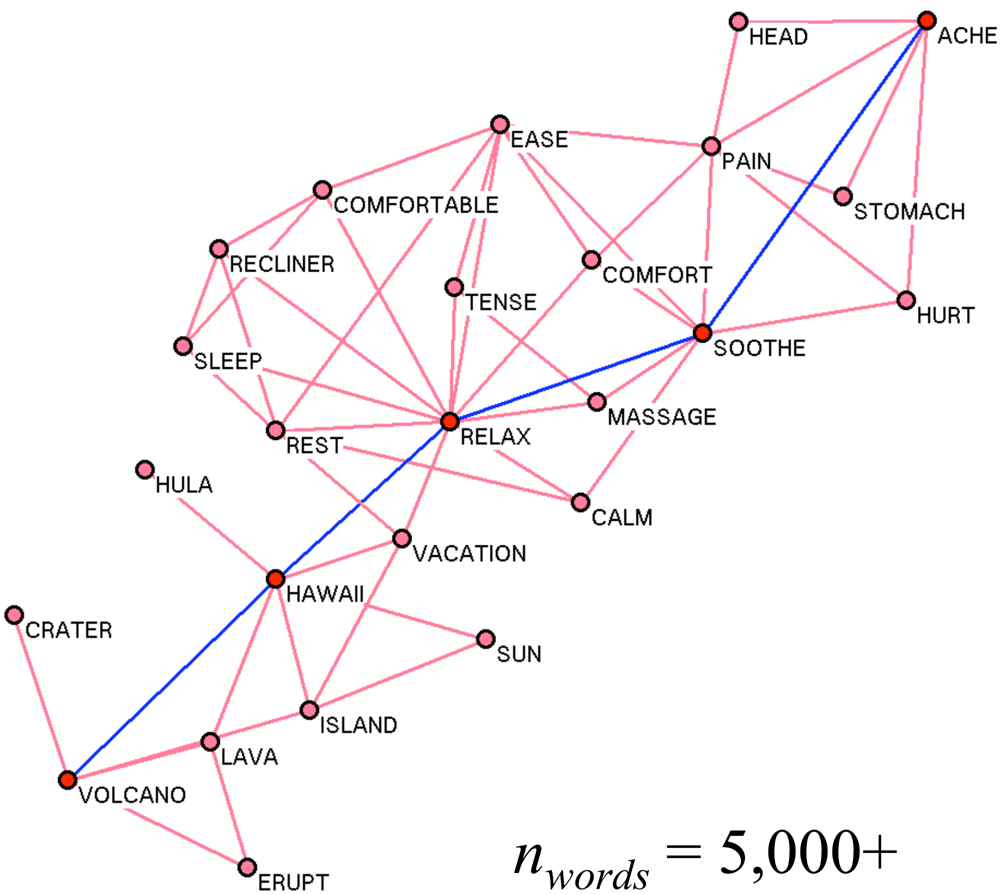
EARTH
PLUTO
JUPITER
NEPTUNE
VENUS
URANUS
SATURN
COMET
MARS
ASTEROID

(Nelson, McEvoy & Schreiber, 1998)

Word association



Word association

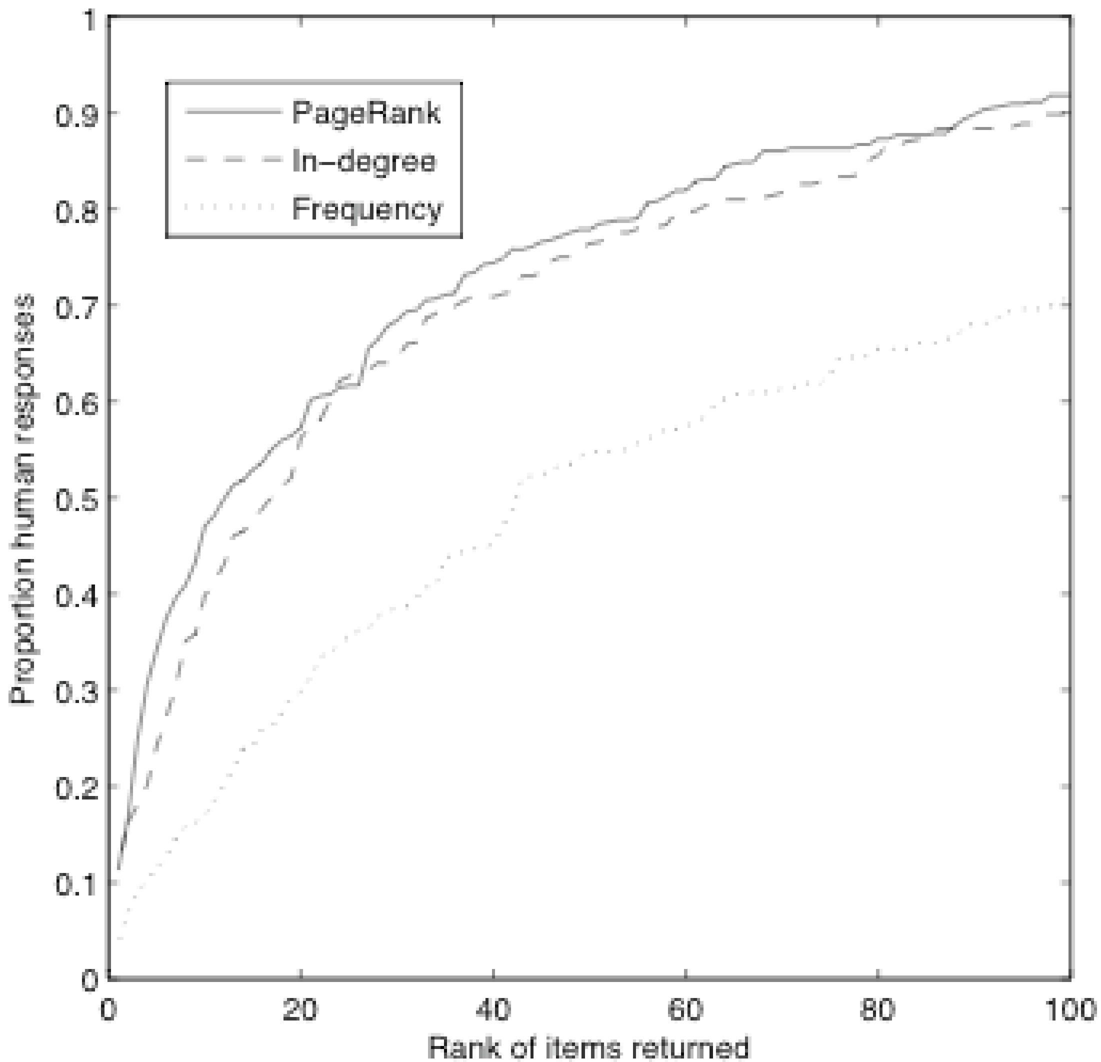


$n_{words} = 5{,}000+$

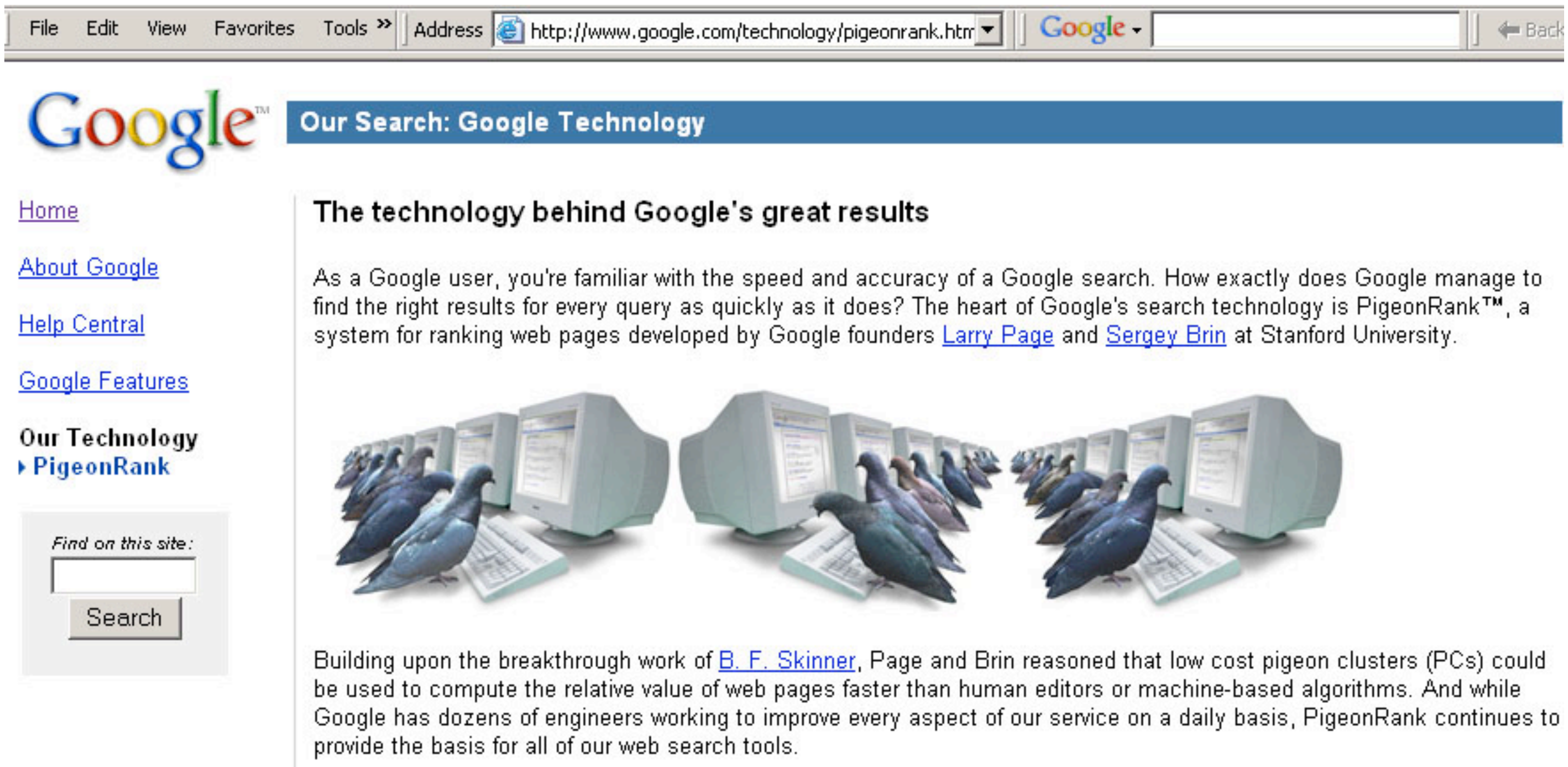
An experiment

- Name the first word that comes into your head beginning with the letter D
- Parallels web search… retrieve a set of words (webpages) that match a letter (query)
- Look at ranks of responses under
  - PageRank of words within word association network
  - number of times word appears as an associate
  - overall word frequency

An experiment



How Google works
(on April 1, 2002)



File Edit View Favorites Tools » Address http://www.google.com/technology/pigeonrank.htm Google Back

Google™

Our Search: Google Technology

[Home](#)

[About Google](#)

[Help Central](#)

[Google Features](#)

Our Technology
▸ PigeonRank

Find on this site:

Search

The technology behind Google's great results

As a Google user, you're familiar with the speed and accuracy of a Google search. How exactly does Google manage to find the right results for every query as quickly as it does? The heart of Google's search technology is PigeonRank™, a system for ranking web pages developed by Google founders [Larry Page](#) and [Sergey Brin](#) at Stanford University.

Building upon the breakthrough work of [B. F. Skinner](#), Page and Brin reasoned that low cost pigeon clusters (PCs) could be used to compute the relative value of web pages faster than human editors or machine-based algorithms. And while Google has dozens of engineers working to improve every aspect of our service on a daily basis, PigeonRank continues to provide the basis for all of our web search tools.

Search and memory

- Human memory and internet search share the problem of retrieving one fact among many
- Under one view of knowledge (semantic networks) the organization of facts is similar
- A simple definition of “importance” works well in both cases…
- Similar correspondences exist for more complex kinds of search (semantic similarity)

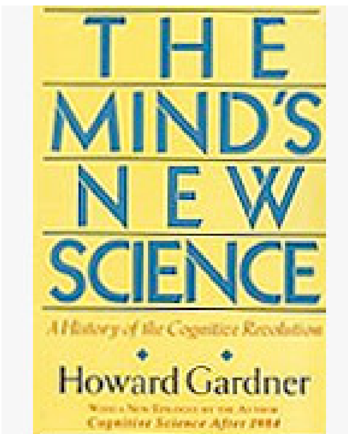
Outline

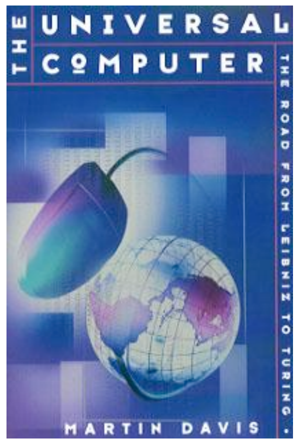
Spam filters and classification

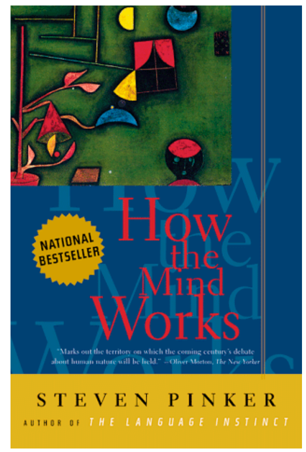
Search and memory

Inferring preferences

Inferring preferences

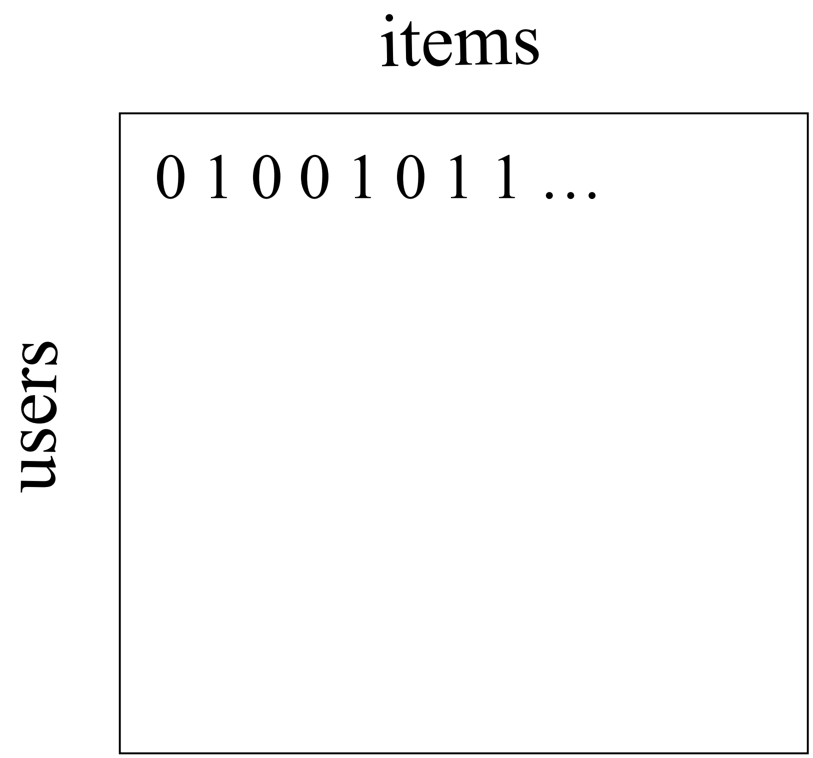






?

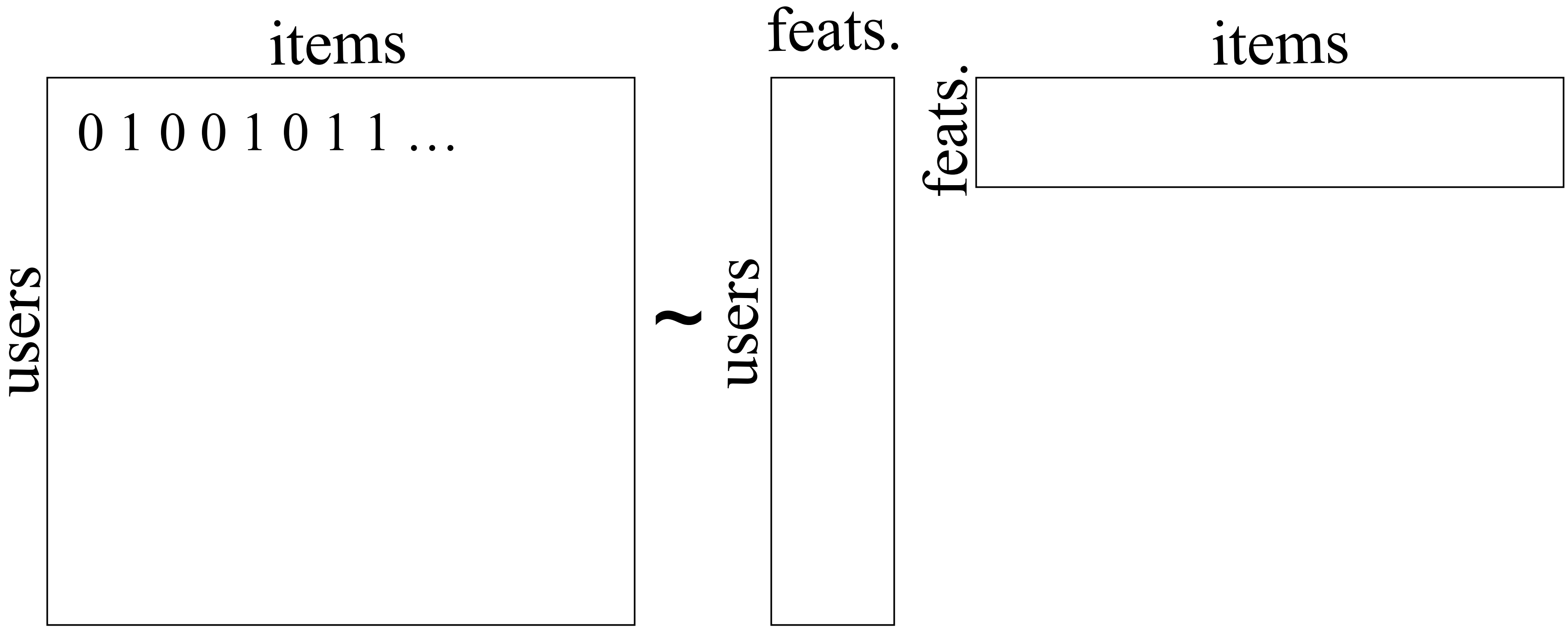
Collaborative filtering



Approaches to collaborative filtering

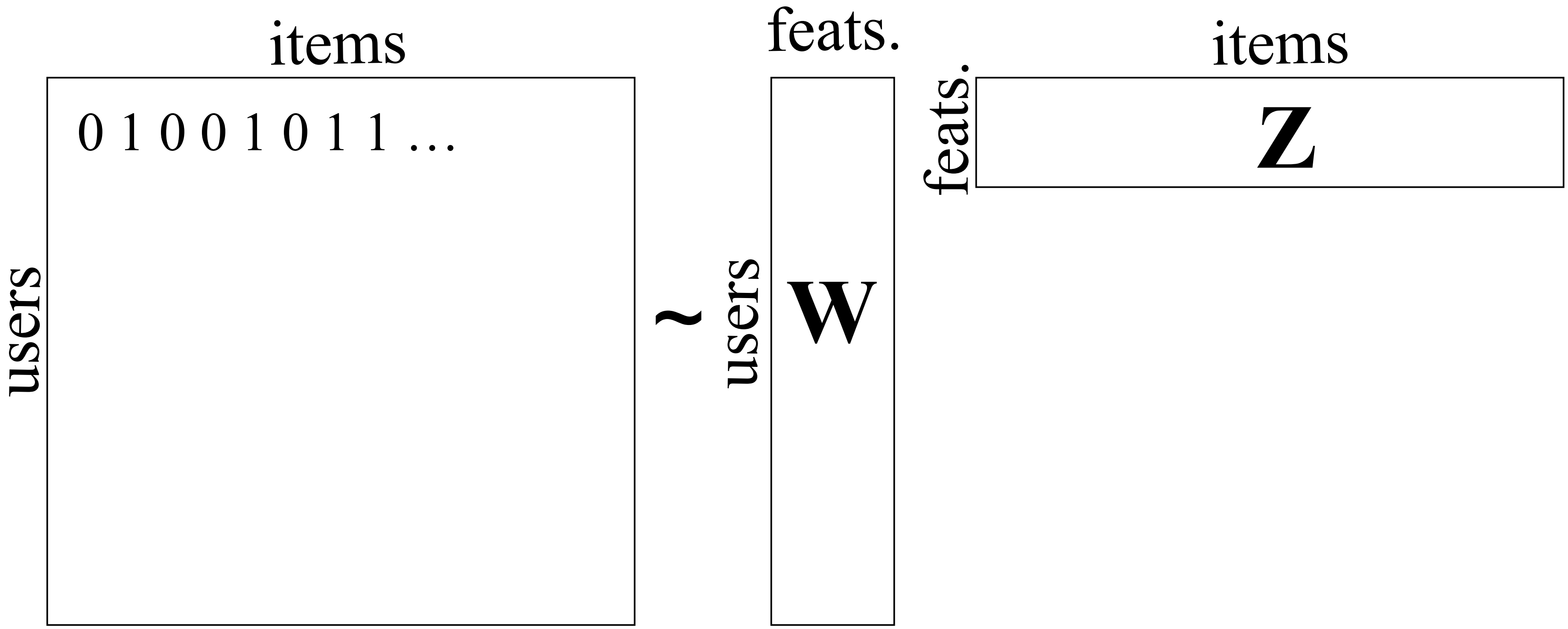
- Simple: compute correlation between users, and use a weighted average of purchases
  - typically divide by item frequency first
- Fast: compute correlation between items
  - can be done quickly when users have few items
- Most expressive: dimensionality reduction
  - make inferences about users and items

Matrix factorization



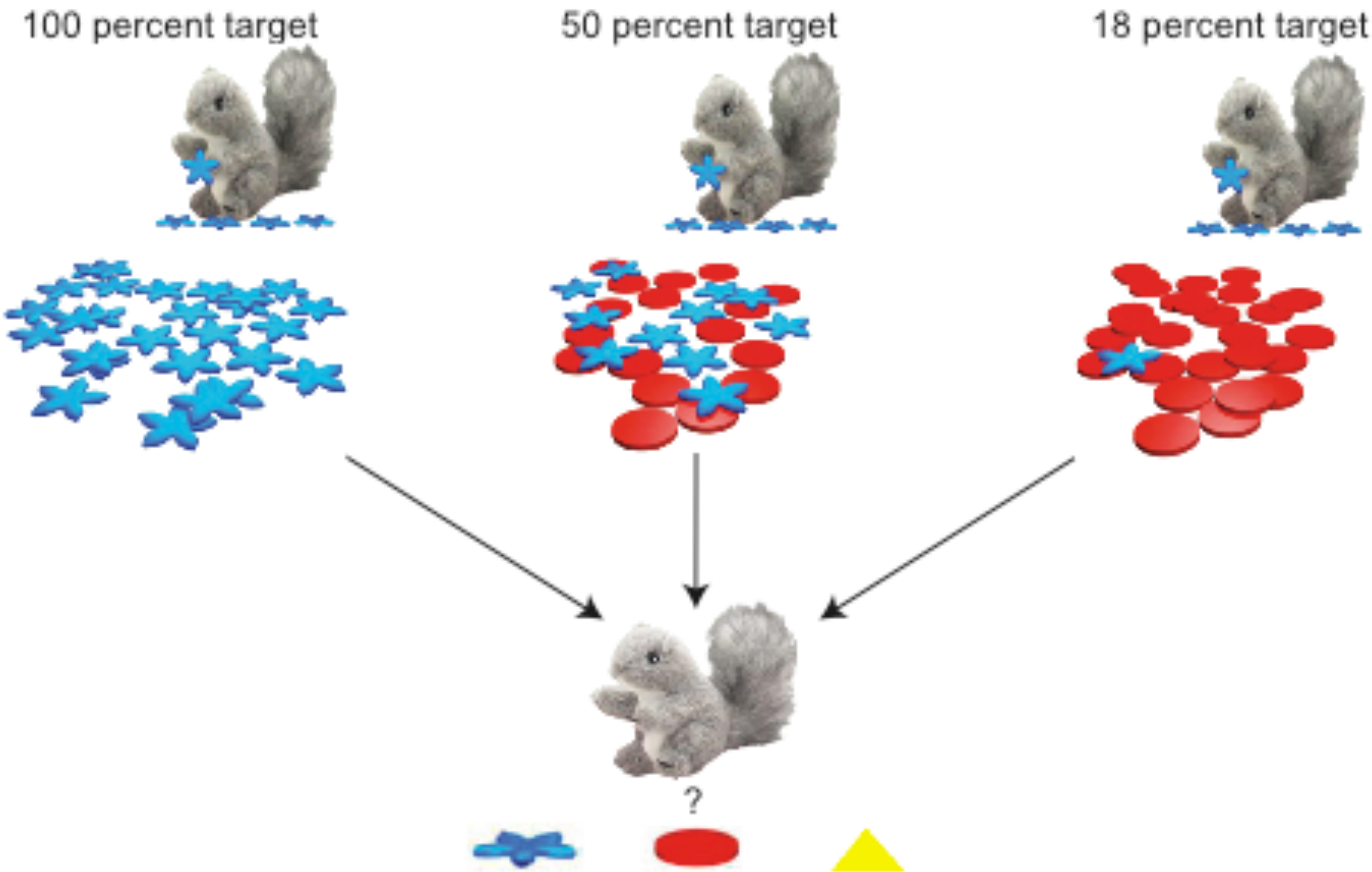
Mixed multinomial logit model

$$p(c = i \mid \mathbf{u}) = \frac{\exp(u_i)}{\sum_j \exp(u_j)} \qquad u_i = \sum_k w_k z_k$$



(McFadden, 1973)

Developing understanding of choice



(Kushnir, Xu & Wellman, 2008)

Relating choice and preference

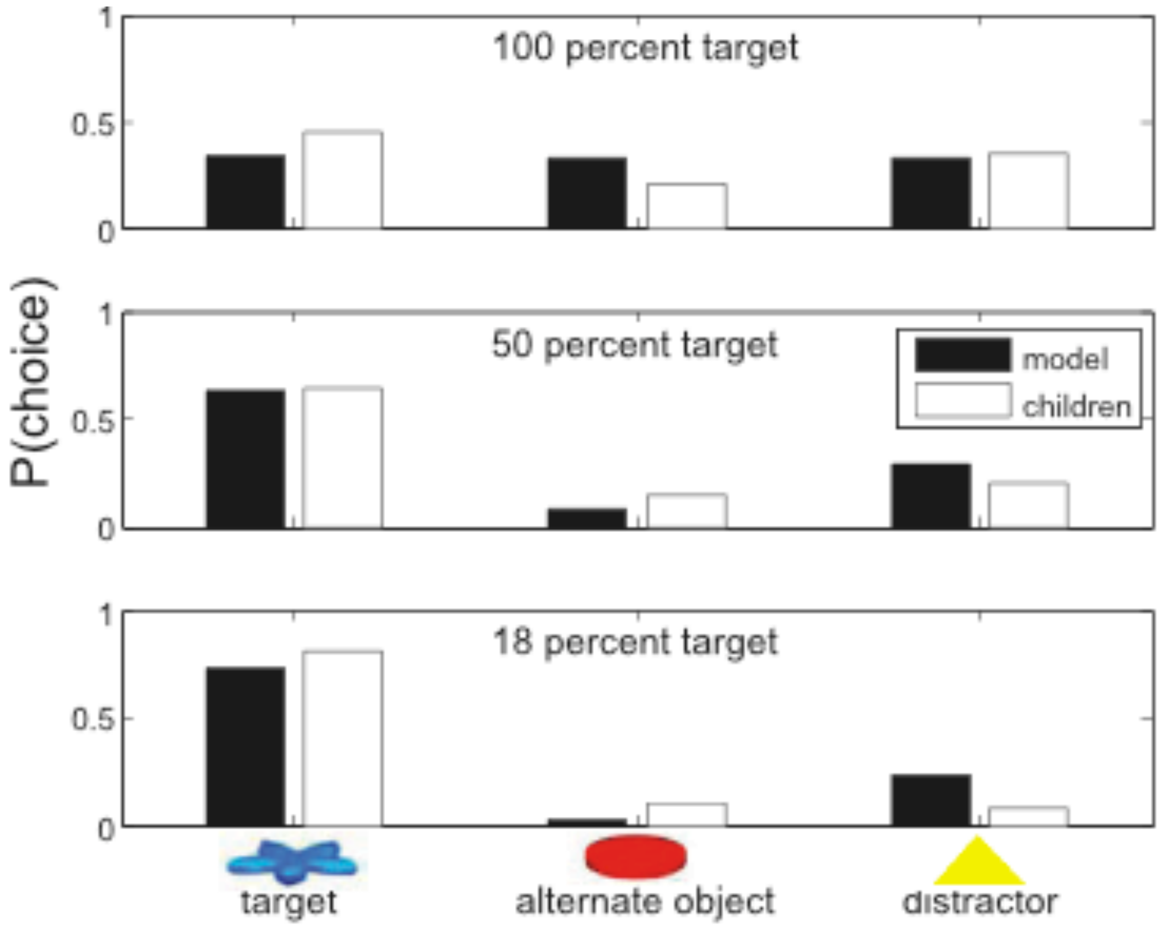
- Assume choices follow the MML model

$$p(c = i \mid \mathbf{u}) = \frac{\exp(u_i)}{\sum_j \exp(u_j)}$$

- Children infer utilities by applying Bayes’ rule

$$p(\mathbf{u} \mid \mathbf{c}) \propto \left[\prod_n p(c_n \mid \mathbf{u})\right] p(\mathbf{u})$$

Developing understanding of choice



(Lucas, Griffiths, Xu & Fawcett, in press)

Inferring preferences

- Collaborative filtering predicts what you will like by using knowledge of what others like
- Different strategies exist, varying in the kinds of information they produce and their runtime
- Even young children are capable of making inferences about preferences, and do so in a way that is consistent with statistical inference

Conclusion

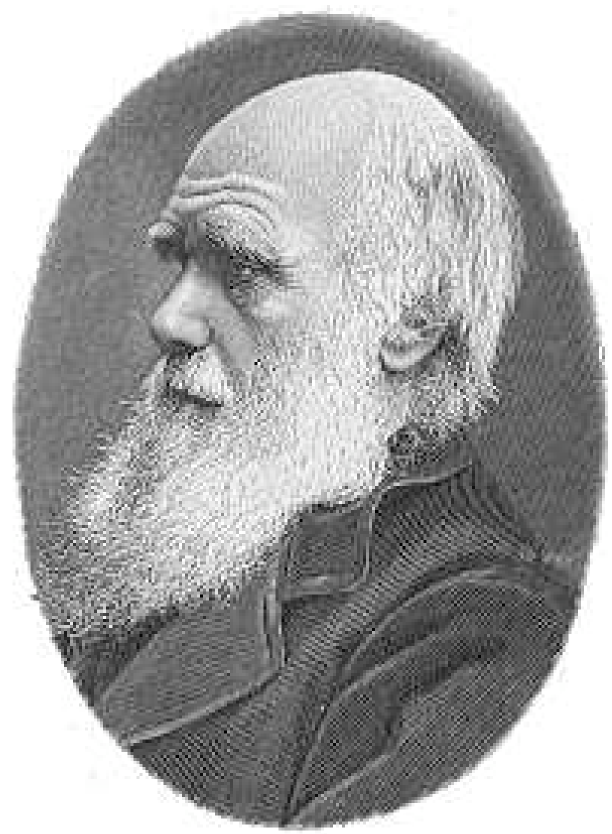
- Brains and computers face similar problems
  - an opportunity for convergent evolution
- We can find connections between the solutions employed by the these systems
- By exploring these connections, we can begin to think about how to help computers solve problems that are currently solved by humans
  - e.g., language learning, causal learning, science

Connecting cultural and biological evolution

Tom Griffiths
Department of Psychology
Cognitive Science Program
University of California, Berkeley

Evolution

- Three key ideas:
  - variation
  - heritable
  - differential reproduction
- Evolution is a theory that naturally lends itself to mathematics…



Charles Darwin

Replicator dynamics

$$\frac{dx_i}{dt} = \sum_j q_{ij} f_j x_j - \phi x_i$$

rate of change of proportion of type *i*

probability type *i* parent has type *j* child

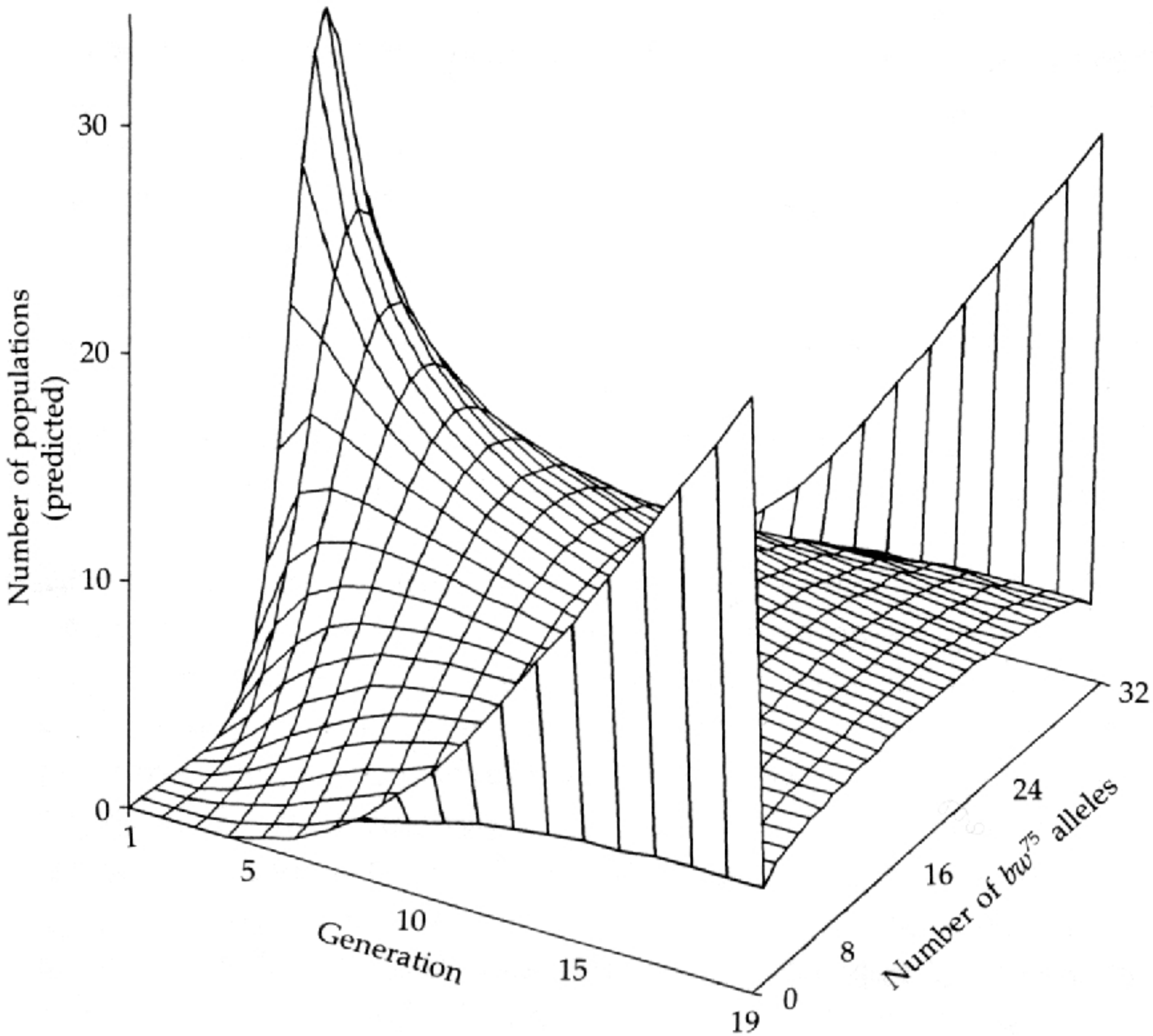
fitness of type *i*

mutation

selection

(no **drift** due to infinite population)

Genetic drift

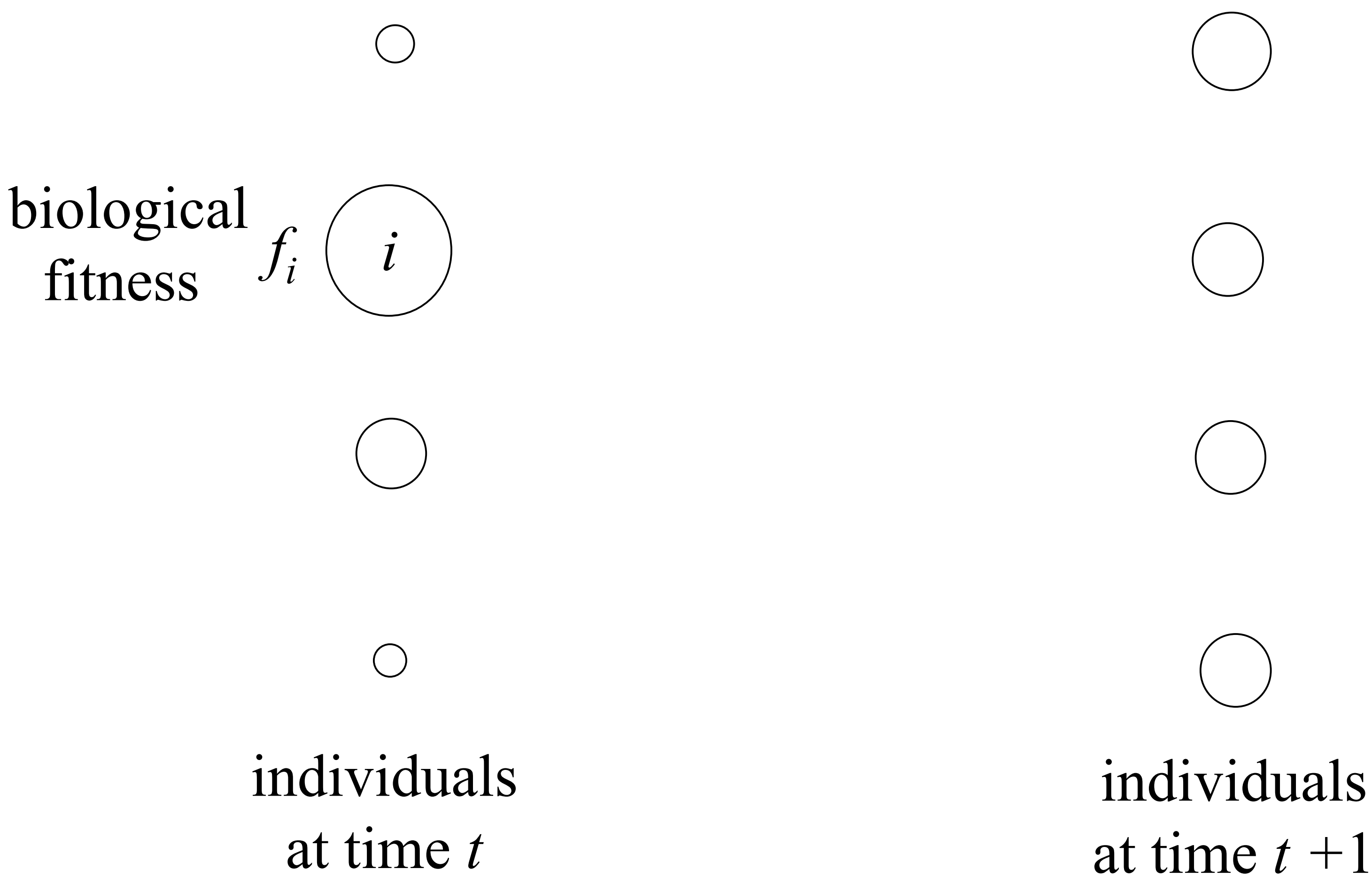


Replicator dynamics

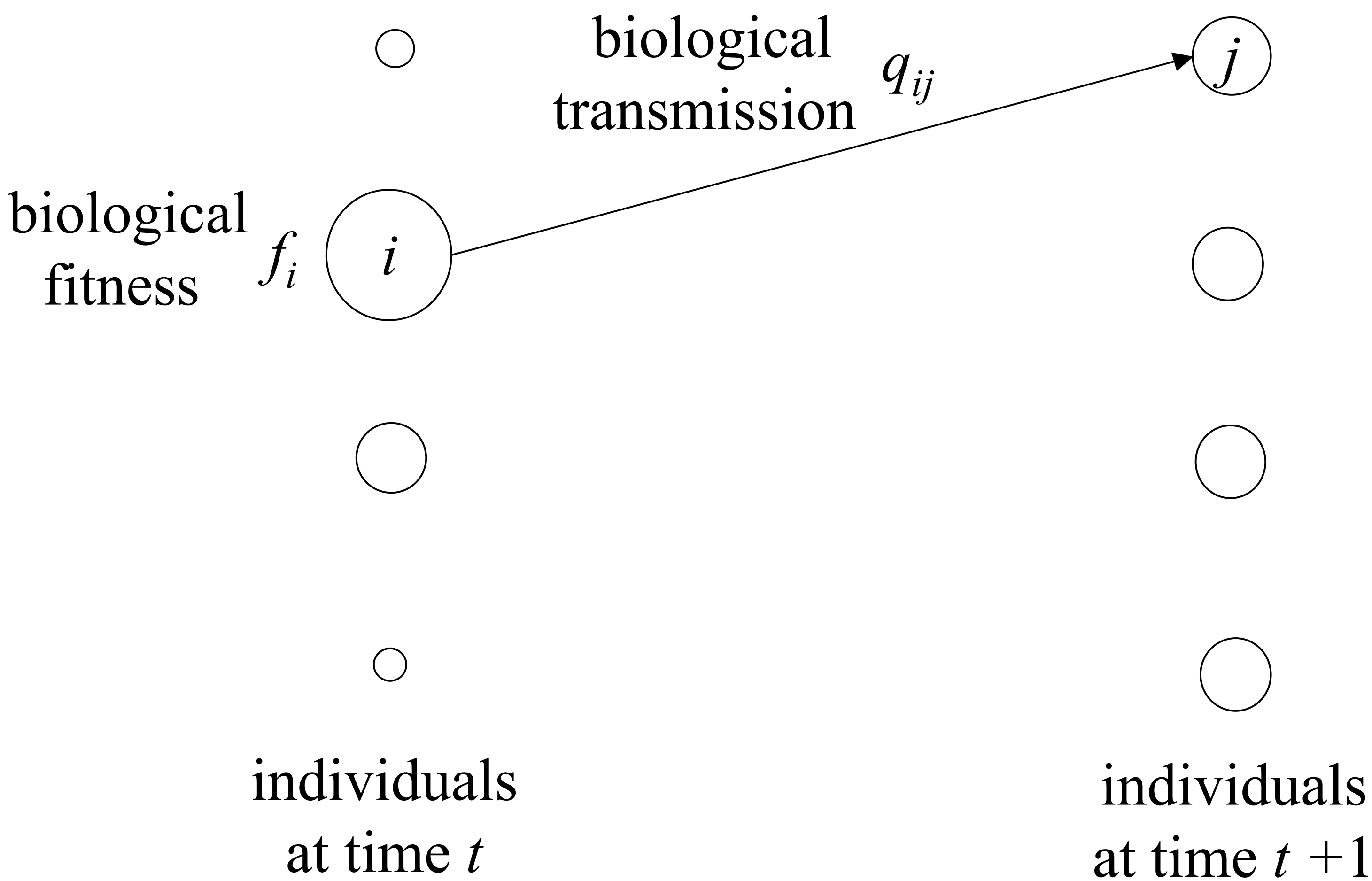
individuals
at time t

individuals
at time $t+1$

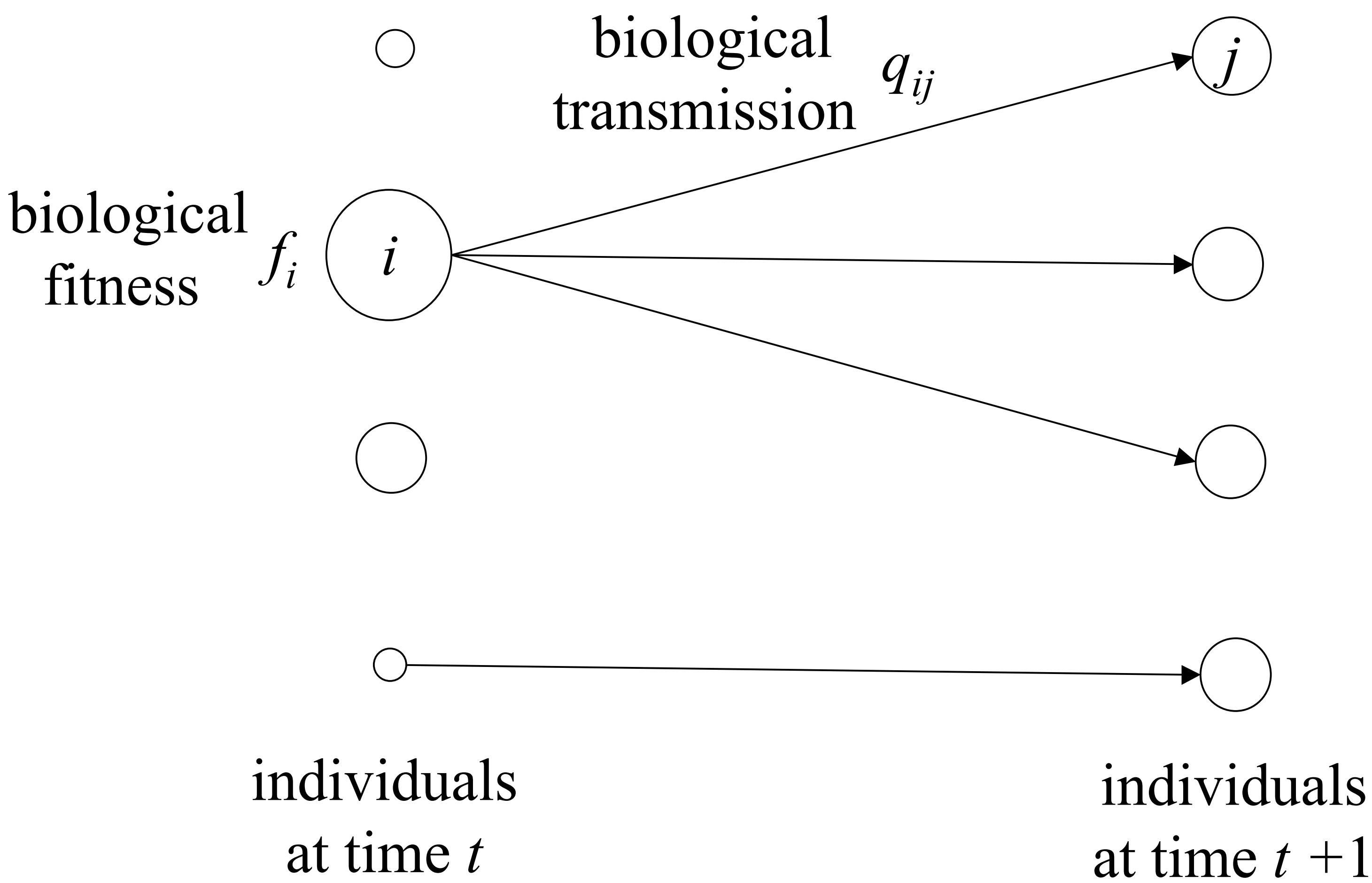
Replicator dynamics



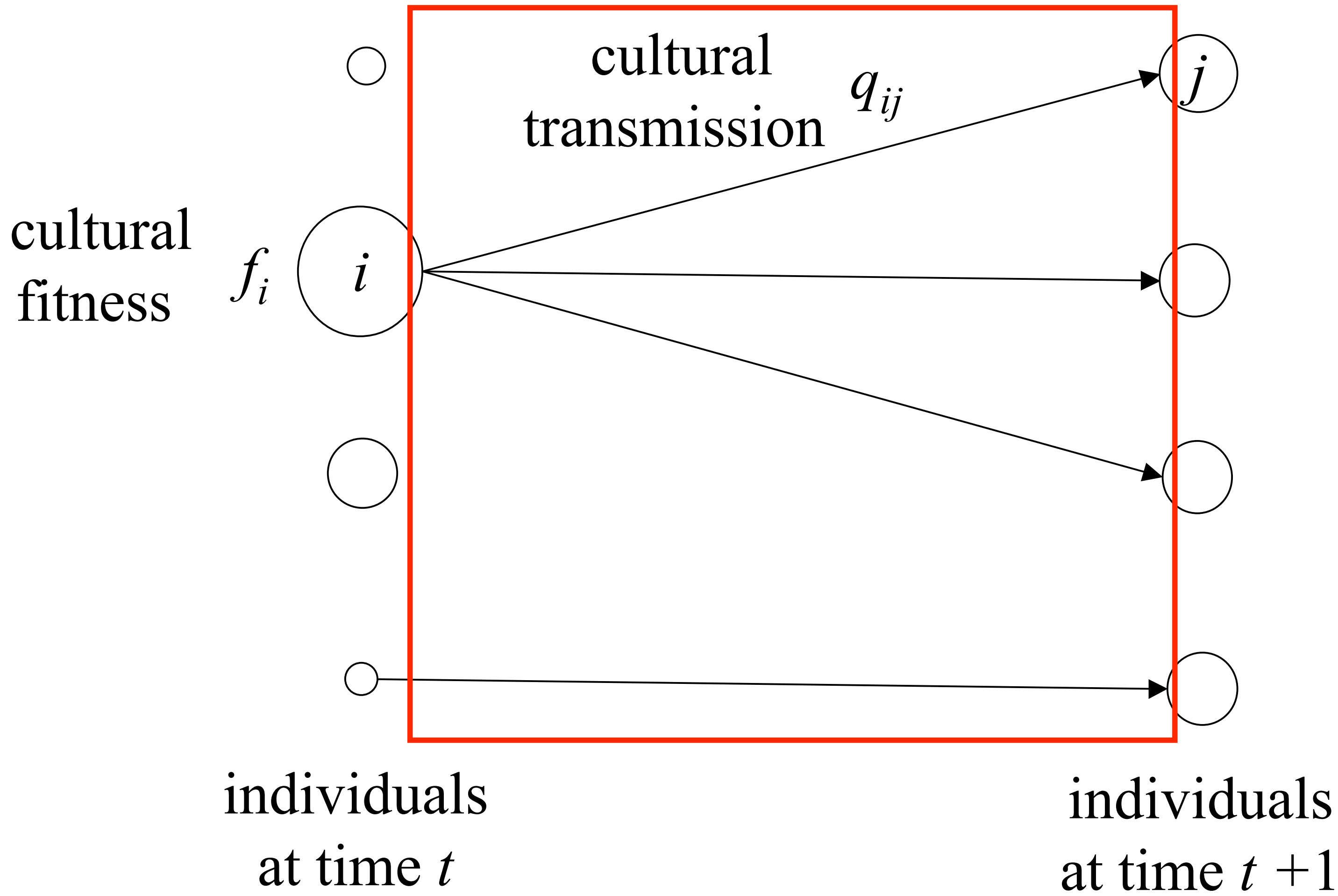
Replicator dynamics



Replicator dynamics



Cultural evolution

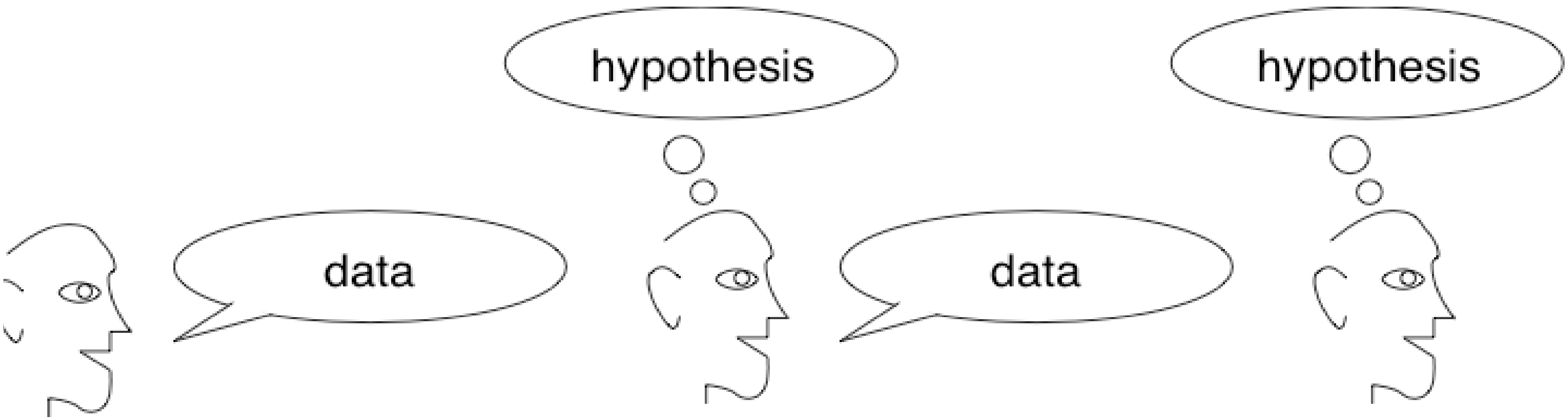


Cultural transmission

How does transmission transform information?

Iterated learning

(Kirby, 2001)



Outline

Part I: Formal analysis of iterated learning

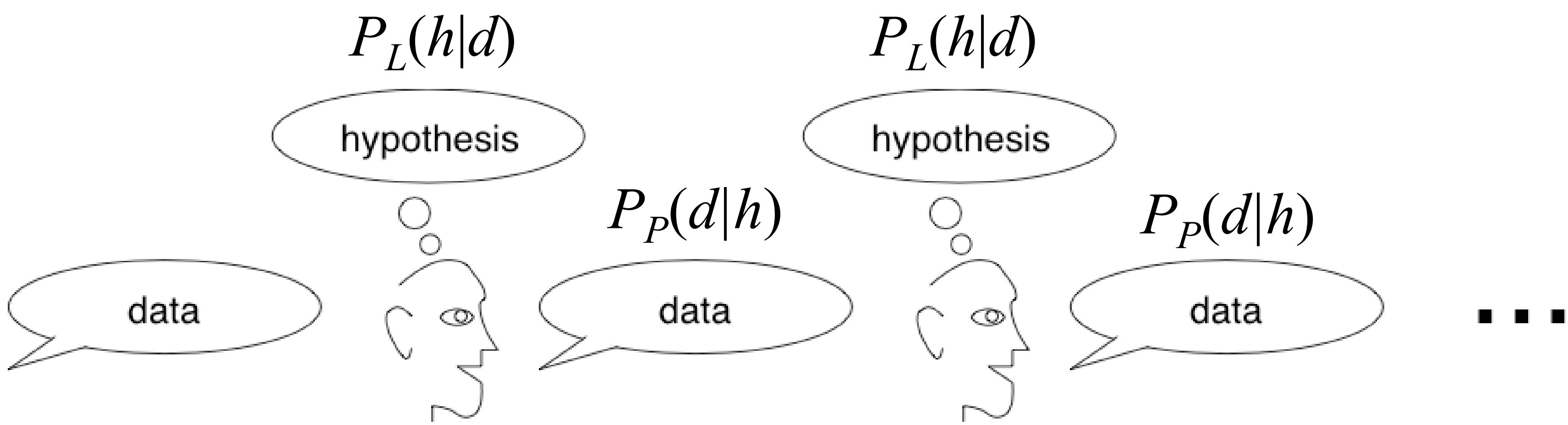
Part II: Iterated learning in the lab

Outline

Part I: Formal analysis of iterated learning

Part II: Iterated learning in the lab

Analyzing iterated learning



$P_L(h|d)$: probability of inferring hypothesis h from data d

$P_P(d|h)$: probability of generating data d from hypothesis h

Markov chains

$$\longrightarrow \boldsymbol{x} \longrightarrow \boldsymbol{x} \longrightarrow \boldsymbol{x} \longrightarrow \boldsymbol{x} \longrightarrow \boldsymbol{x} \longrightarrow \boldsymbol{x} \longrightarrow \boldsymbol{x} \longrightarrow \boldsymbol{x} \longrightarrow$$

Transition matrix

$q_{ij}=P(\boldsymbol{x}^{(t+1)}=j|\boldsymbol{x}^{(t)}=i)$

- Variables $\boldsymbol{x}^{(t+1)}$ independent of history given $\boldsymbol{x}^{(t)}$
- Converges to a *stationary distribution* under easily checked conditions (i.e., if it is ergodic)

Analyzing iterated learning

$$d_0 \xrightarrow[P_L(h|d)]{} h_1 \xrightarrow[P_P(d|h)]{} d_1 \xrightarrow[P_L(h|d)]{} h_2 \xrightarrow[P_P(d|h)]{} d_2 \xrightarrow[P_L(h|d)]{} h_3 \text{ —}$$

A Markov chain on hypotheses

$$h_1 \xrightarrow[\underbrace{\Sigma_d P_P(d|h) P_L(h|d)}_{\text{corresponds to } q_{ij} \text{ in replicator dynamics}}]{} h_2 \xrightarrow[\Sigma_d P_P(d|h) P_L(h|d)]{} h_3 \text{ —}$$

Bayesian inference

Reverend Thomas Bayes

Bayes' theorem

Posterior probability

Likelihood

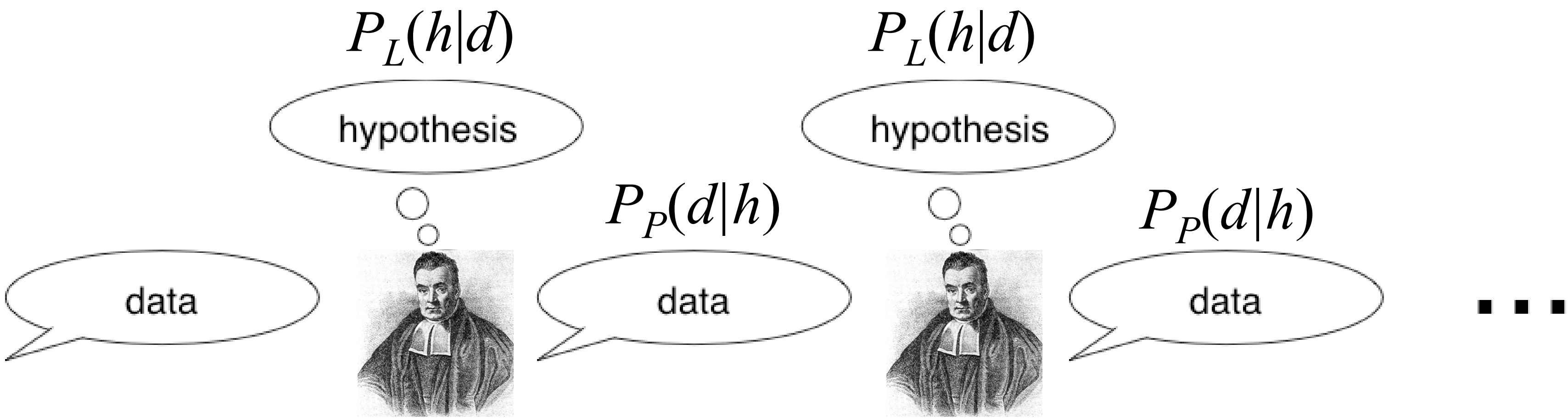
Prior probability

$$P(h \mid d) = \frac{P(d \mid h)P(h)}{\sum_{h' \in H} P(d \mid h')P(h')}$$

Sum over space of hypotheses

h: hypothesis
d: data

Iterated Bayesian learning



Assume learners *sample* from their posterior distribution:

$$P_L(h \mid d) = \frac{P_P(d \mid h)P(h)}{\sum_{h' \in H} P_P(d \mid h')P(h')}$$

Stationary distributions

- Markov chain on h converges to the prior, $P(h)$
  - the probability of choosing a hypothesis converges to the prior probability of that hypothesis

- Intuitively,each inference allows the prior to affect the hypothesis chosen, with the prior itself being the only distribution not modified

(Griffiths & Kalish, 2005)

Back to the replicator dynamics…

- Replicator dynamics

$$\frac{dx_i}{dt} = \sum_j q_{ij} f_j x_j - \phi x_i$$

- "Neutral model" (f_j constant)

$$\frac{dx_i}{dt} = \sum_j q_{ij} x_j - x_i \qquad \frac{d\mathbf{x}}{dt} = (\mathbf{Q} - \mathbf{I})\mathbf{x}$$

- Stable equilibrium at first eigenvector of **Q**, which is our stationary distribution

Analyzing iterated learning

- The outcome of iterated learning is strongly affected by the inductive biases of the learners
  - hypotheses with high prior probability ultimately appear with high probability in the population
- Establishes a connection between constraints on learning and cultural universals…
- …and provides formal justification for the idea that culture reflects the structure of the mind

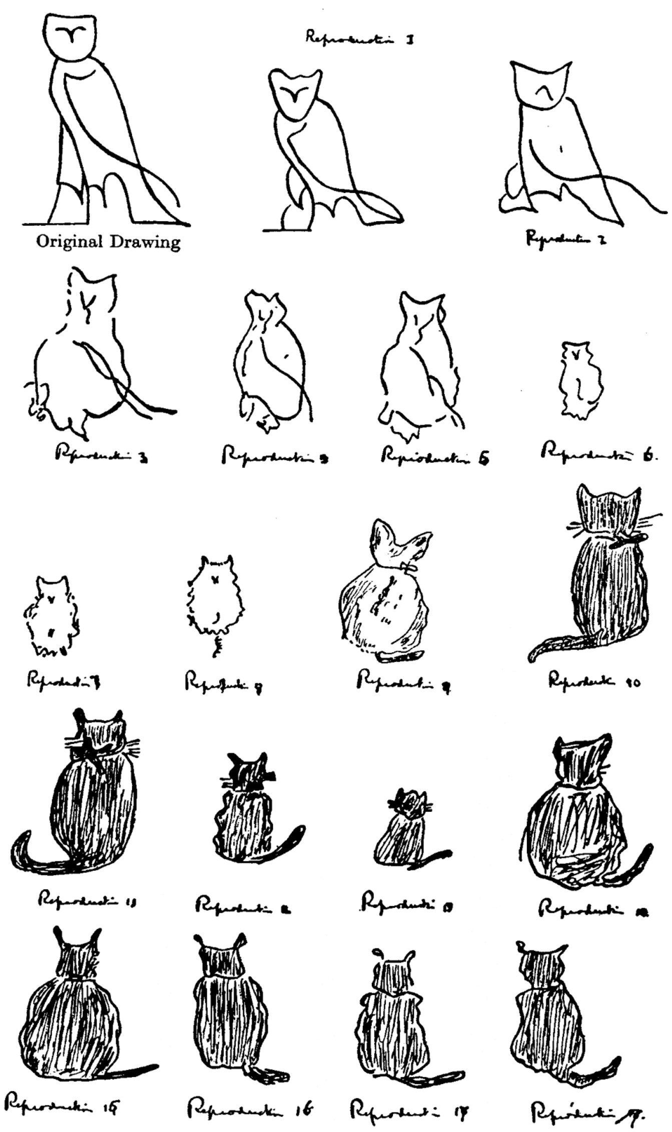
Outline

Part I: Formal analysis of iterated learning

Part II: Iterated learning in the lab

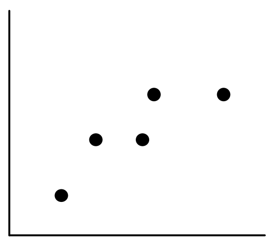
Serial reproduction

(Bartlett, 1932)

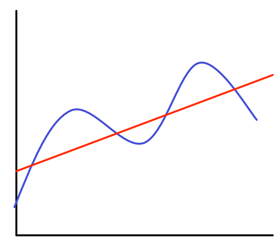


Iterated function learning

data



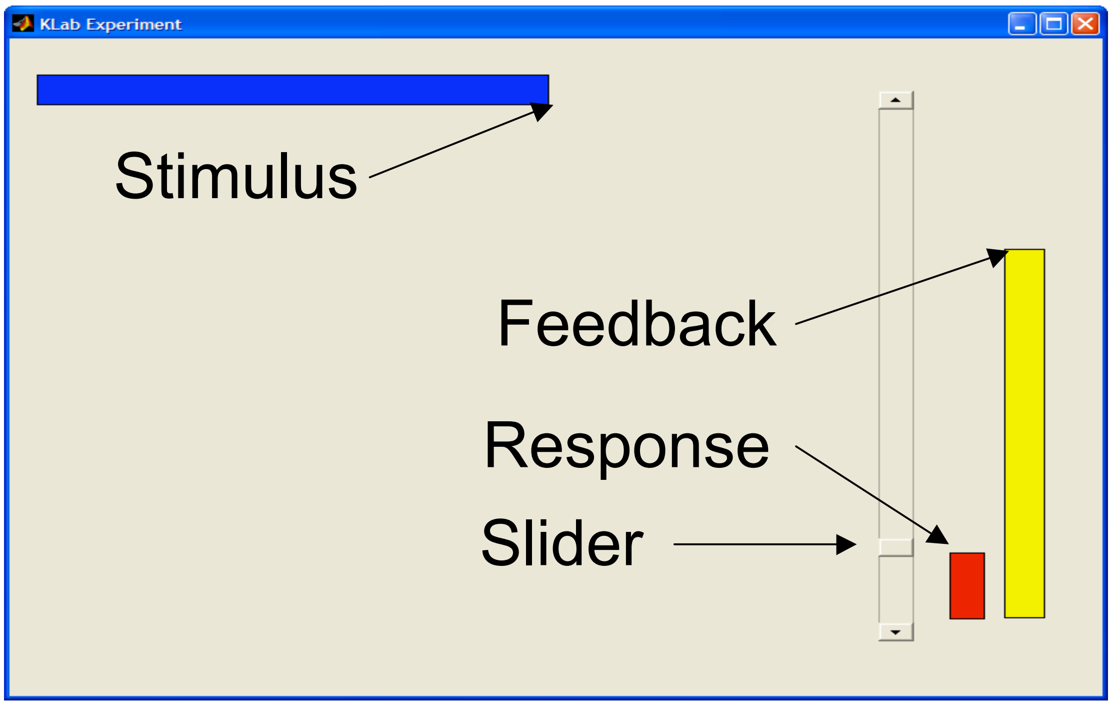
hypotheses



- Each learner sees a set of (*x*,*y*) pairs
- Makes predictions of *y* for new *x* values
- Predictions are data for the next learner

(Kalish, Griffiths, & Lewandowsky, 2007)

Function learning experiments



Examine iterated learning with different initial data

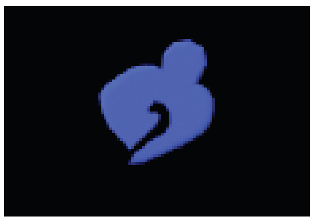
Initial data

Iteration

1 2 3 4 5 6 7 8 9

Frequency distributions

data

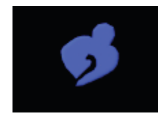


5 x "DUP"

5 x "NEK"

(Vouloumanos, 2008)

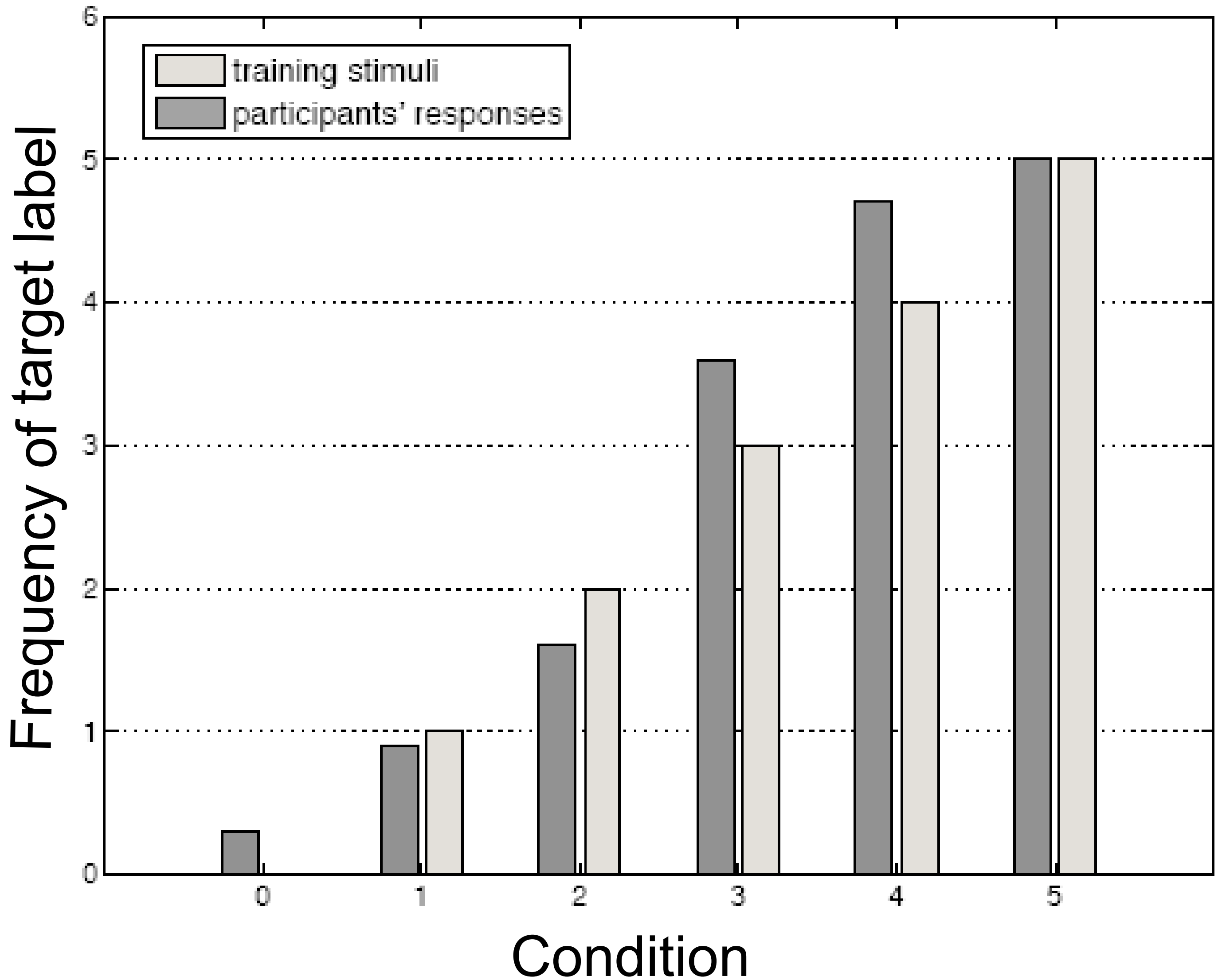
hypotheses

$P($"DUP"$|$  $) = \theta$

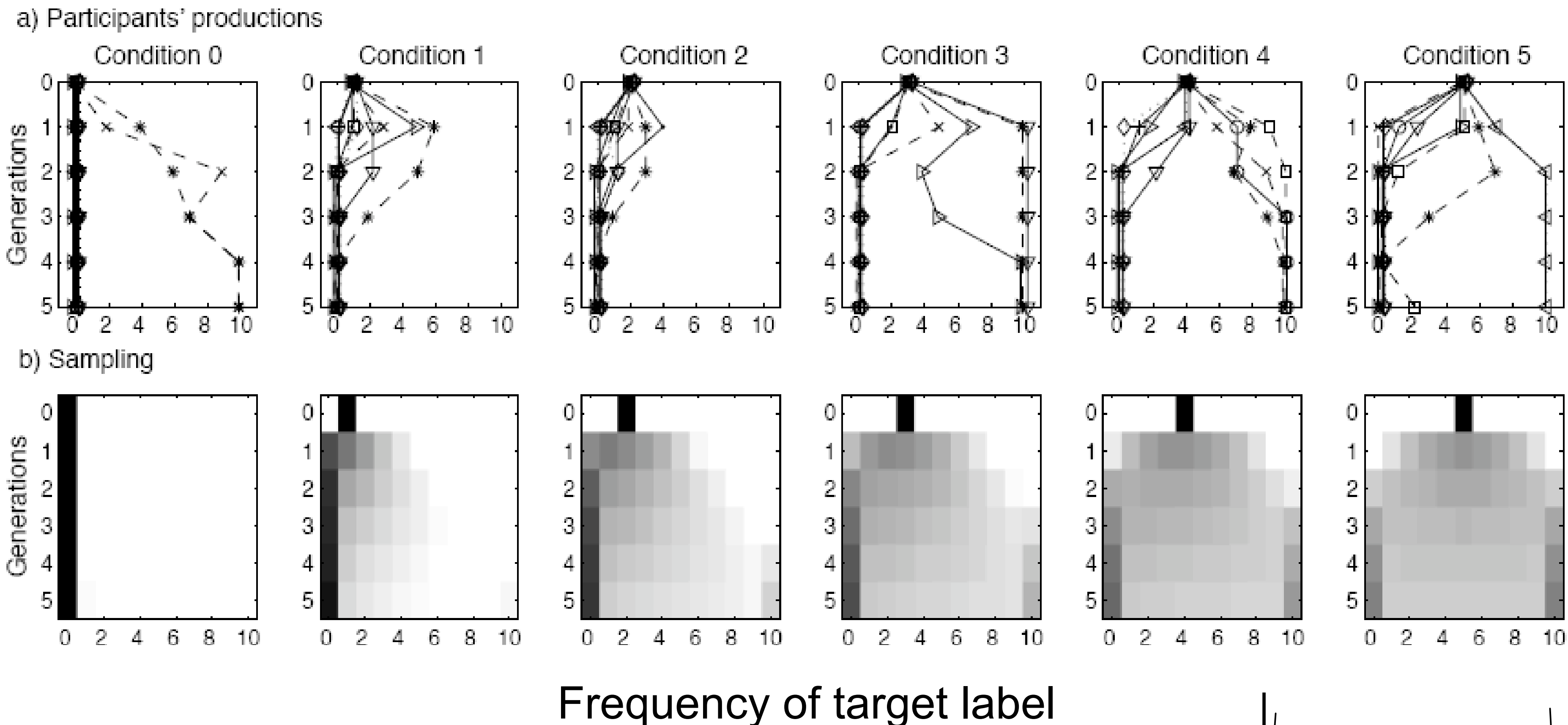
- Each learner sees objects receiving two labels
- Produces labels for those objects at test
- First learner: one label $\{0,1,2,3,4,5\}/10$ times

(Reali & Griffiths, in press)

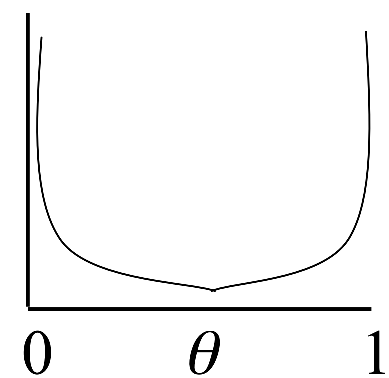
Results after one generation



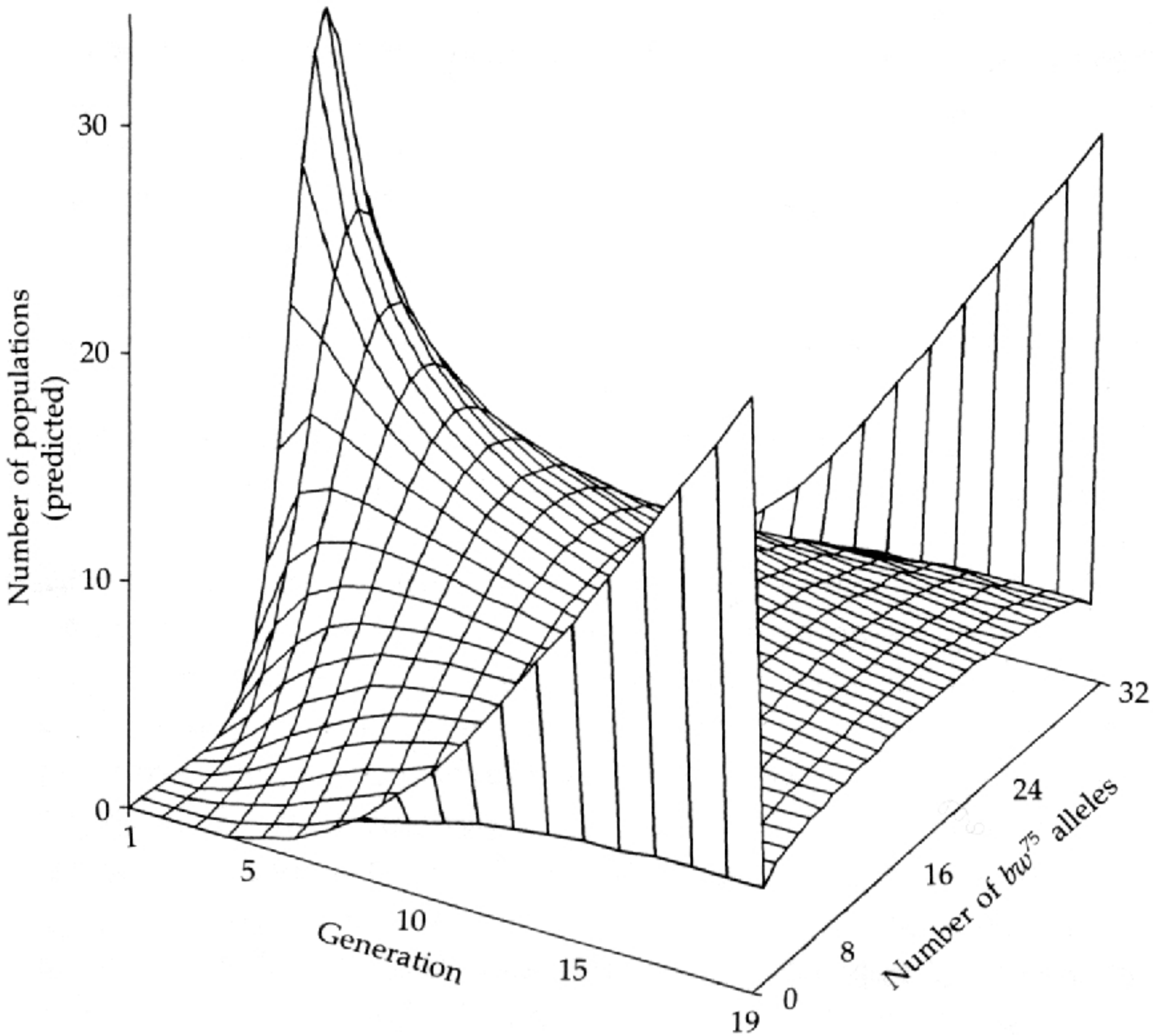
Results after five generations



Bayesian model has a prior favoring regularization:



Genetic drift



Conclusions

- Cultural transmission can systematically alter information being transmitted
- The result of iterated learning is strongly influenced by constraints on learning
- Despite different mechanisms, formal analogies exist between biological and cultural evolution
  - learning = mutation (but is a directed process)
  - drift = drift (and can be a useful explanatory tool)