The Language Organism
Lecture 7: Cultural evolution by iterated learning

Simon Kirby

simon@ling.ed.ac.uk

Learning bias

Learning bias

- Particular properties that a learner has will influence that learner's ability to acquire particular languages

- This *learning bias* comes from our biology (right?), and has been argued to affect the languages that we speak

Learning bias

- Particular properties that a learner has will influence that learner's ability to acquire particular languages
- This *learning bias* comes from our biology (right?), and has been argued to affect the languages that we speak
- Christiansen & Devlin (1997) suggest that sequential learning biases make certain word-orders more easy to learn than others
- For example: verb-final languages are far more likely to have postpositions than verb-initial languages

Learning bias

- Particular properties that a learner has will influence that learner's ability to acquire particular languages
- This *learning bias* comes from our biology (right?), and has been argued to affect the languages that we speak
- Christiansen & Devlin (1997) suggest that sequential learning biases make certain word-orders more easy to learn than others
- For example: verb-final languages are far more likely to have postpositions than verb-initial languages
- Explained by fact that a neural network finds the rare orders harder to learn (tested by feeding particular hand-constructed languages to the neural network)

But what’s missing here?

- In Christiansen & Devlin’s work, the experimenter provides a language, and tests the network’s ability to learn it

- Two issues:

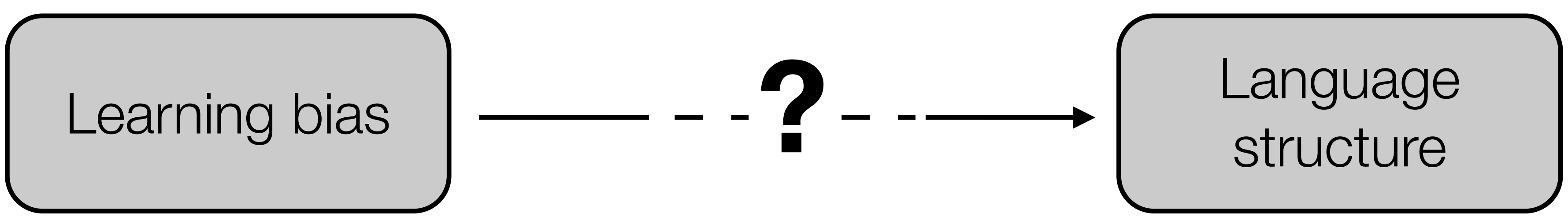
But what’s missing here?

- In Christiansen & Devlin’s work, the experimenter provides a language, and tests the network’s ability to learn it

- Two issues:
    - Where does this language come from?

But what's missing here?

- In Christiansen & Devlin's work, the experimenter provides a language, and tests the network's ability to learn it
- Two issues:
  - Where does this language come from?
  - How do we bridge the gap between learning bias and universal properties of language structure?



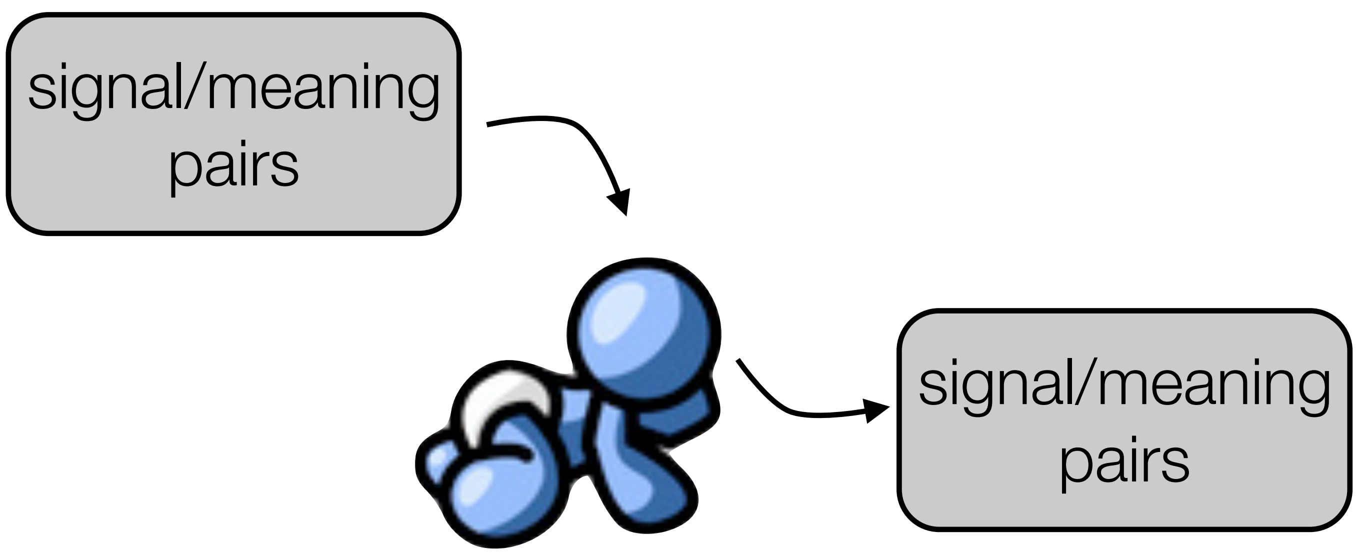
THE PROBLEM OF LINKAGE

Solving the problem of linkage

- Where does the language data come from that our learners have to acquire?

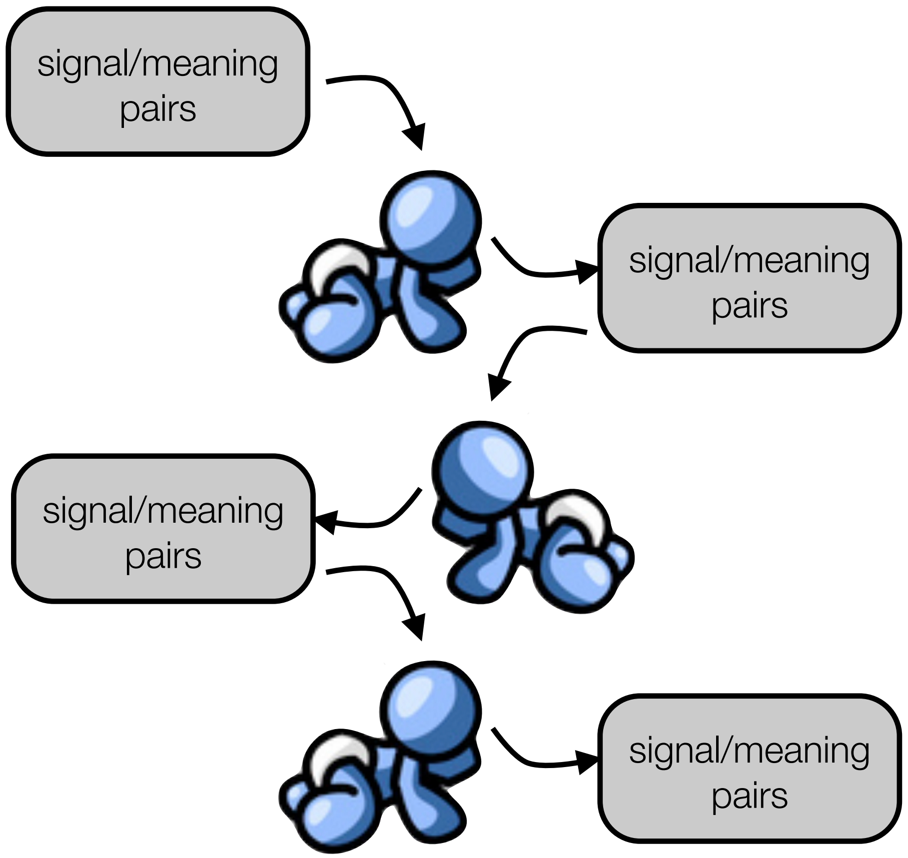
Solving the problem of linkage

- Where does the language data come from that our learners have to acquire?



Solving the problem of linkage

- Where does the language data come from that our learners have to acquire?
- From other learners!



Solving the problem of linkage

- Where does the language data come from that our learners have to acquire?

- From other learners!

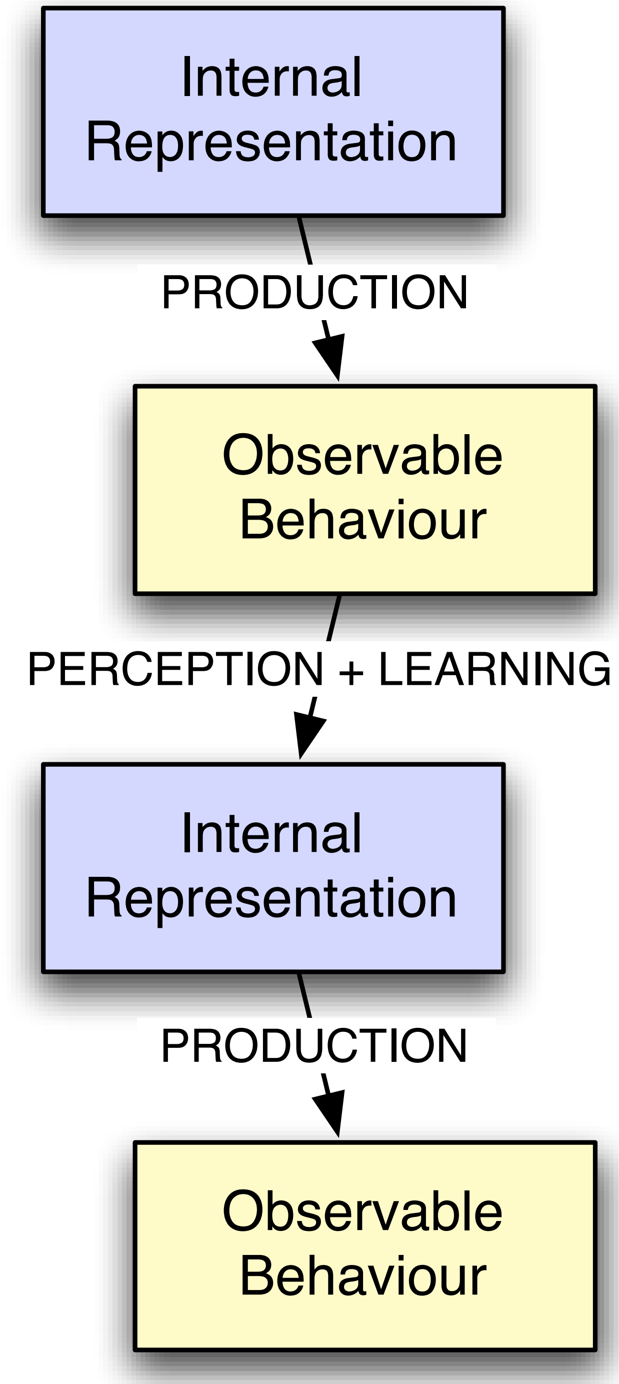
- Language persists over time by repeatedly being learned and used by multiple individuals in a population

- It is out of this continual process of *iterated learning* that the structure of language emerges

- Note, this is *cultural* rather than biological evolution

Solving the problem of linkage

- Where does the language data come from that our learners have to acquire?
- From other learners!
- Language persists over time by repeatedly being learned and used by multiple individuals in a population
- It is out of this continual process of *iterated learning* that the structure of language emerges
- Note, this is *cultural* rather than biological evolution



Solving the problem of linkage

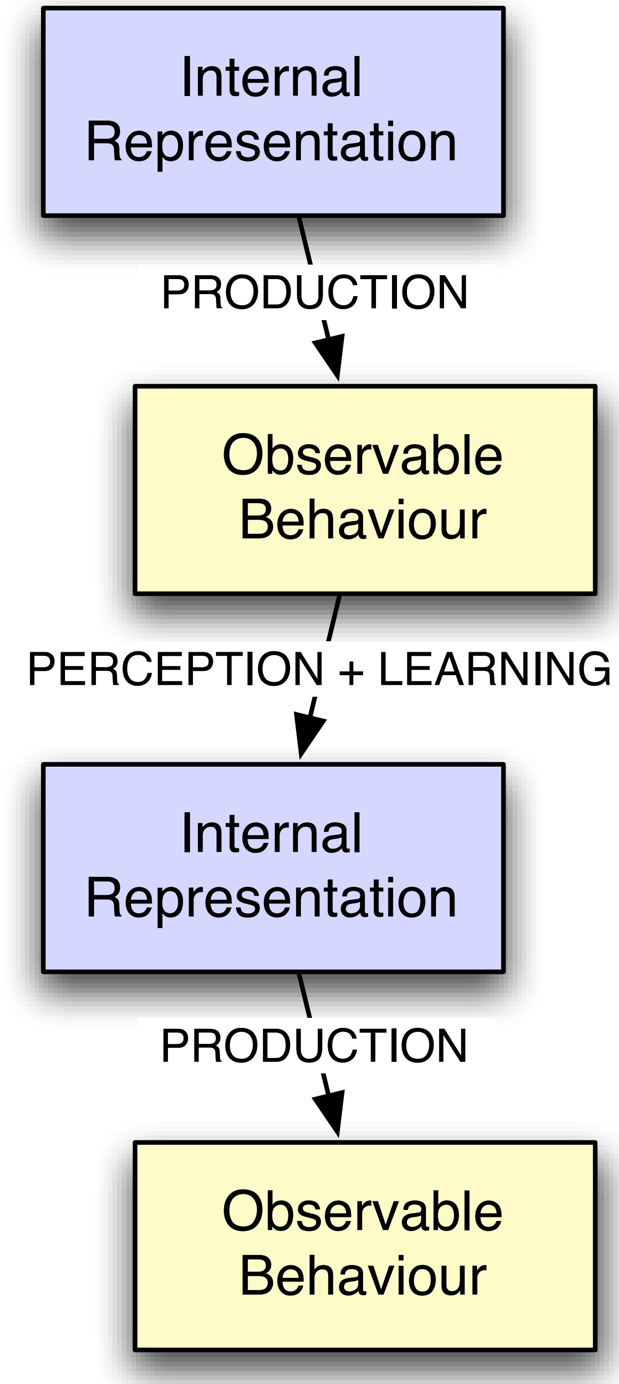
- Where does the language data come from that our learners have to acquire?

- From other learners!

- Language persists over time by repeatedly being learned and used by multiple individuals in a population

- It is out of this continual process of *iterated learning* that the structure of language emerges

- Note, this is *cultural* rather than biological evolution



Key research question for iterated learning

What is the relationship between learning bias and emergent universal properties of language structure?

Key research question for iterated learning

What is the relationship between learning bias and emergent universal properties of language structure?

- We can try and tackle this by using our computational model
- Place agents in population in which they learn from each others utterances.
- Start with random language and observe what languages emerge given different possible learning rules.

Back to our learner

- Idea was to see how different learning biases have distinct consequences in iterated learning even for a hugely simplified model of language
- Varying α, β, γ and δ has surprising consequences

Back to our learner

- Idea was to see how different learning biases have distinct consequences in iterated learning even for a hugely simplified model of language
- Varying α, β, γ and δ has surprising consequences
- Smith (2002) suggests three important tests:

  Acquisition
  Maintenance
  Construction

The acquisition test

- Similar to Christiansen & Devlin’s (1997) approach. Test a learner with particular bias and see if it can learn

The acquisition test

- Similar to Christiansen & Devlin's (1997) approach. Test a learner with particular bias and see if it can learn

- But what type of language?

The acquisition test

- Similar to Christiansen & Devlin's (1997) approach. Test a learner with particular bias and see if it can learn
- But what type of language?
- Try the optimal system (i.e. one that is unambiguous)
- For example, given rule [1, 0, 0, 0]:
  - take a learner using that rule
  - expose them to perfect system
  - test if they can reproduce the perfect system

Result

Result

- Can they reproduce a perfect system? **YES.**

Result

- Can they reproduce a perfect system? **YES.**
- If you use this rule:

$\Delta w_{m_i s_j}$ = **+1** if both m_i and s_j are active
$\Delta w_{m_i s_j}$ = **0** if m_i is active and s_j is inactive
$\Delta w_{m_i s_j}$ = **0** if m_i is inactive and s_j is active
$\Delta w_{m_i s_j}$ = **0** if both m_i and s_j are inactive

- You can successfully learn the perfect system with enough exposure

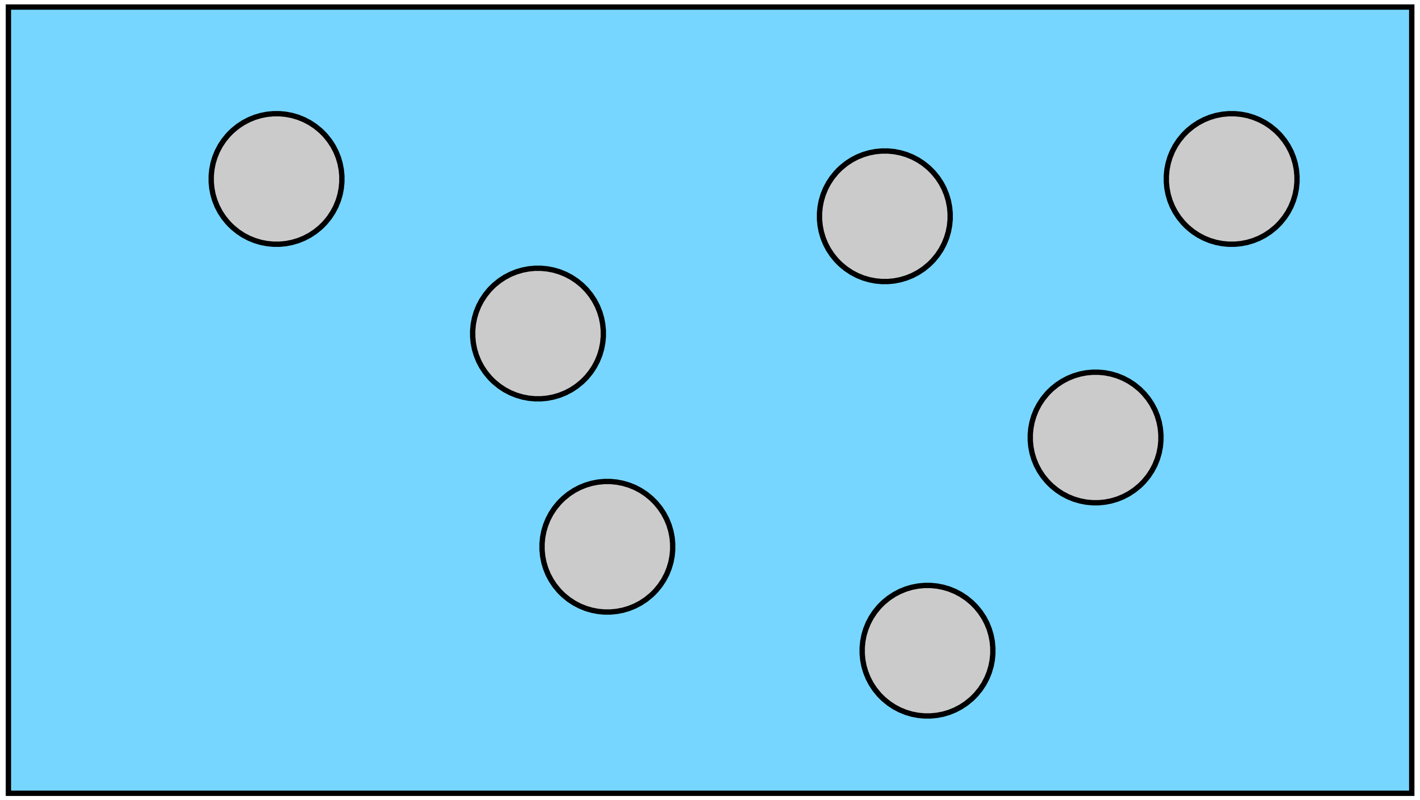
Result

- Can they reproduce a perfect system? **YES.**
- If you use this rule:

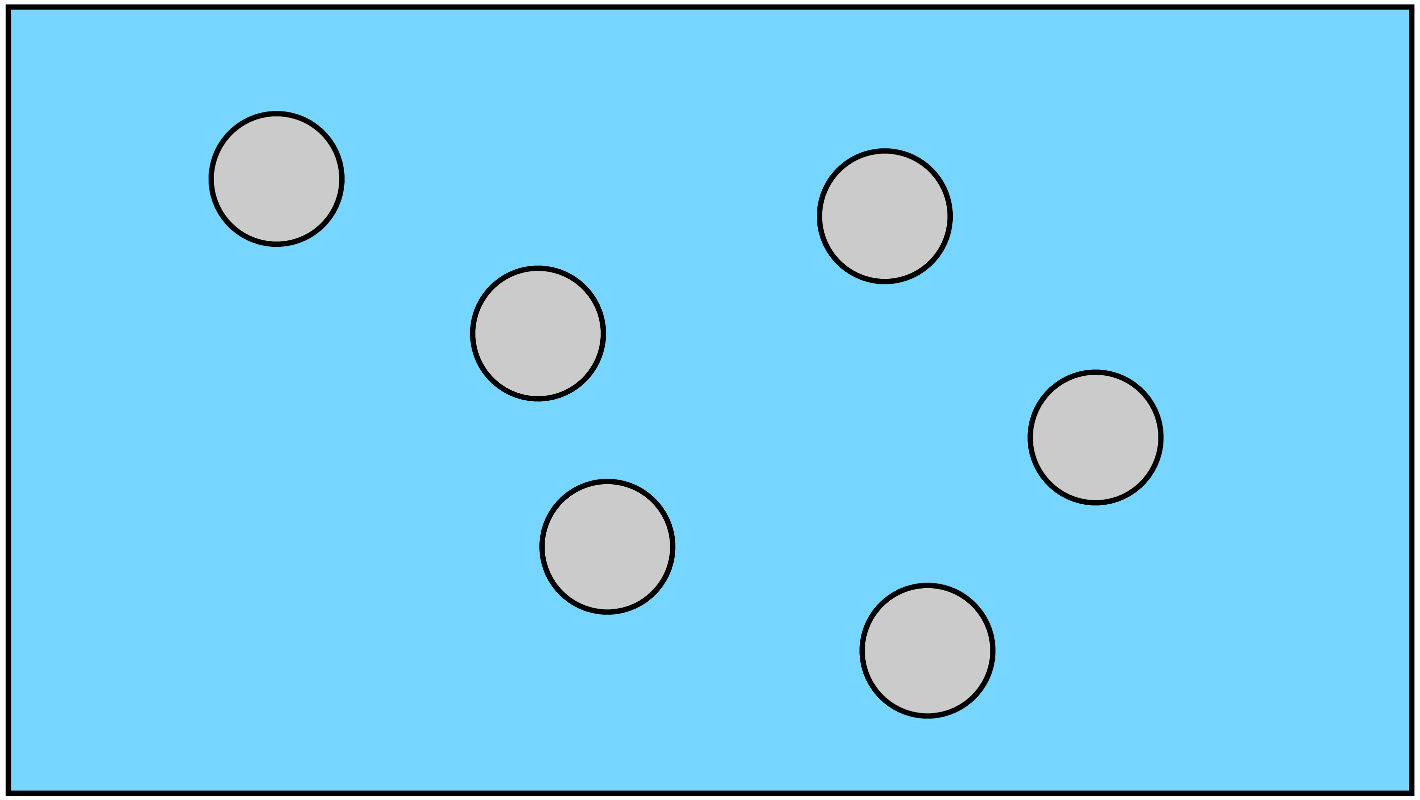
$\Delta w_{m_i s_j}$ = **+1** if both m_i and s_j are active
$\Delta w_{m_i s_j}$ = **0** if m_i is active and s_j is inactive
$\Delta w_{m_i s_j}$ = **0** if m_i is inactive and s_j is active
$\Delta w_{m_i s_j}$ = **0** if both m_i and s_j are inactive

- You can successfully learn the perfect system with enough exposure
- This tells us about the link between learning bias and language learning... but it still doesn't solve the problem of linkage.
- We need to build an *iterated learning model*

Population model

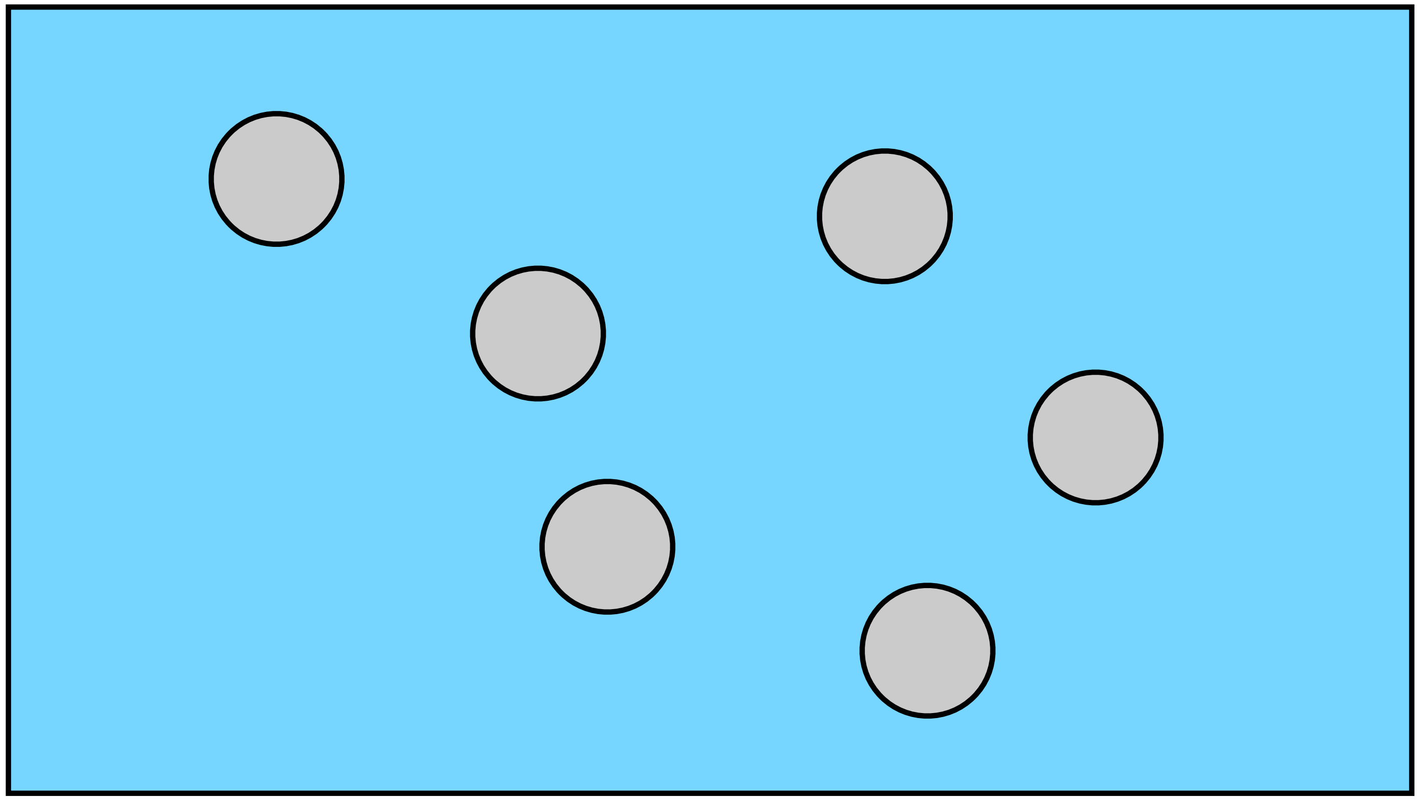


Population model



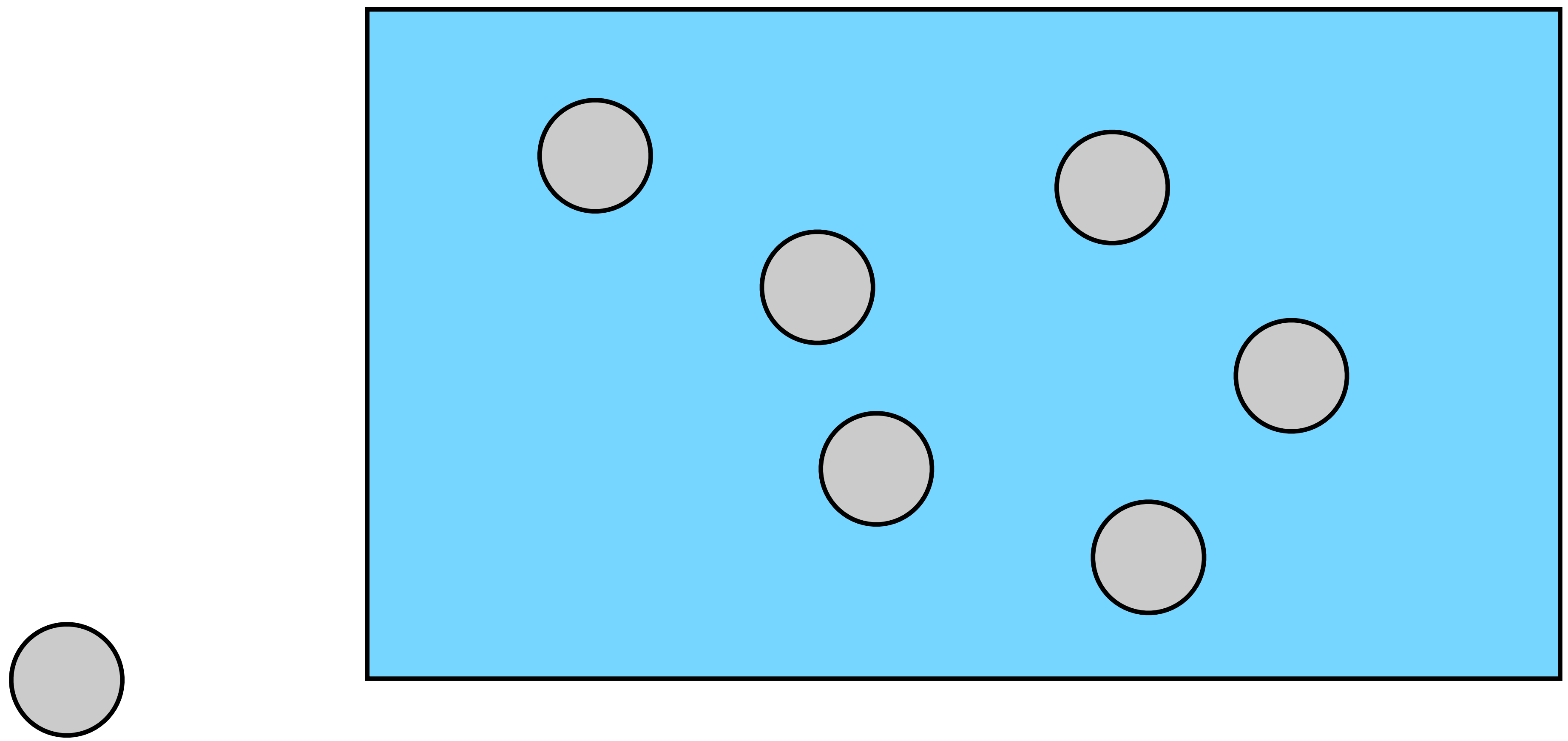
1. Somebody dies

Population model



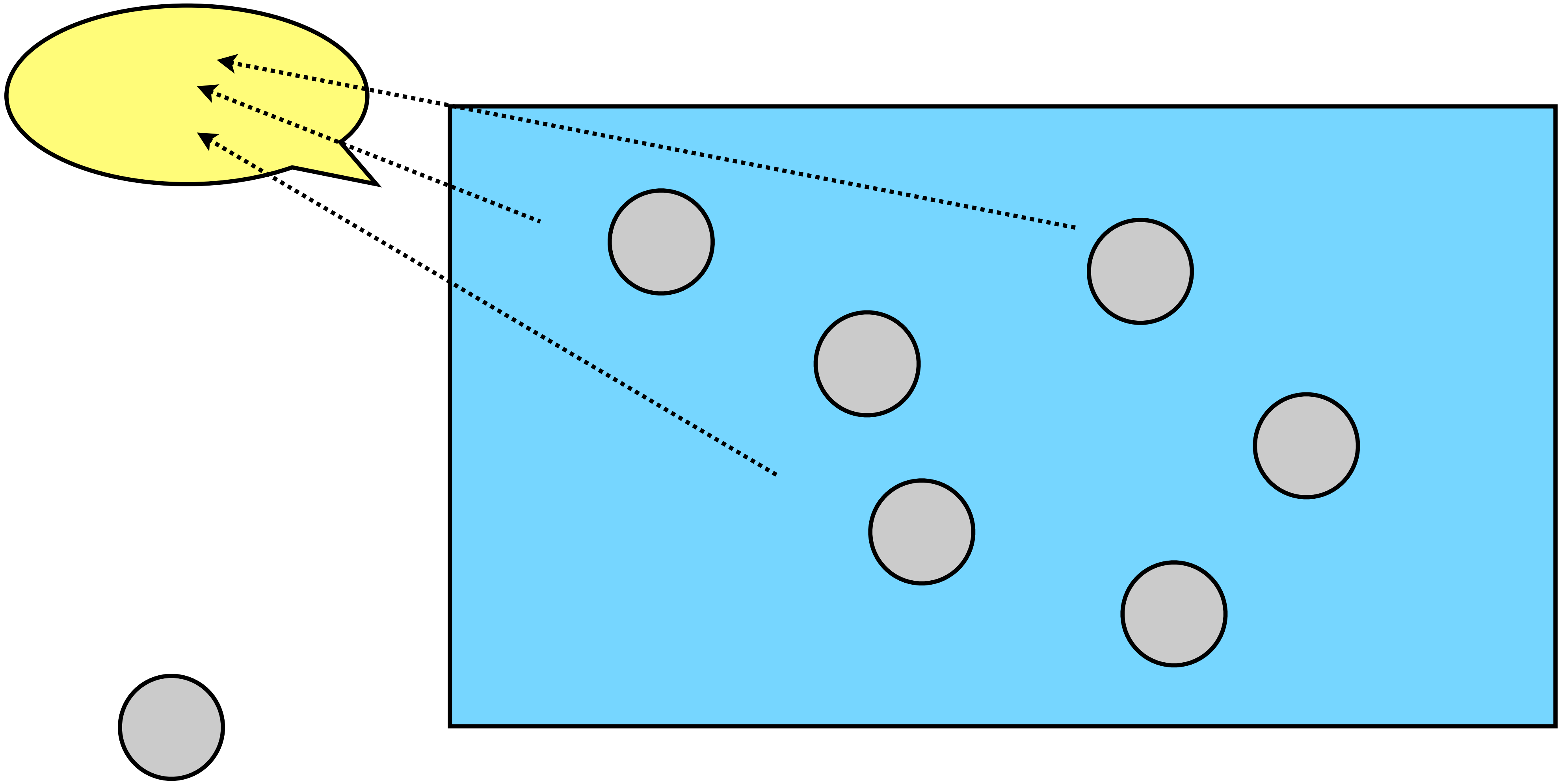
2. A child is born

Population model



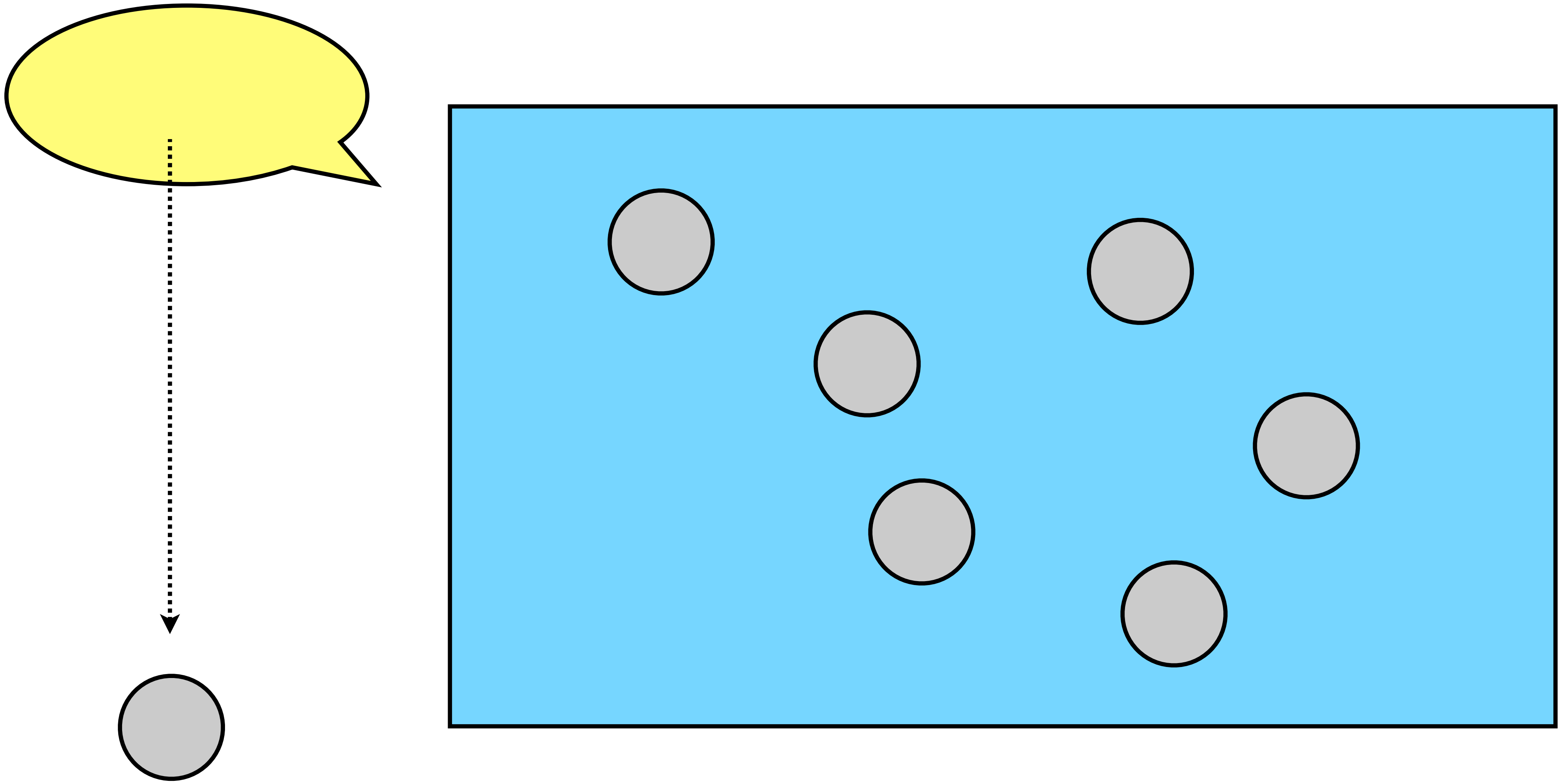
2. A child is born

Population model



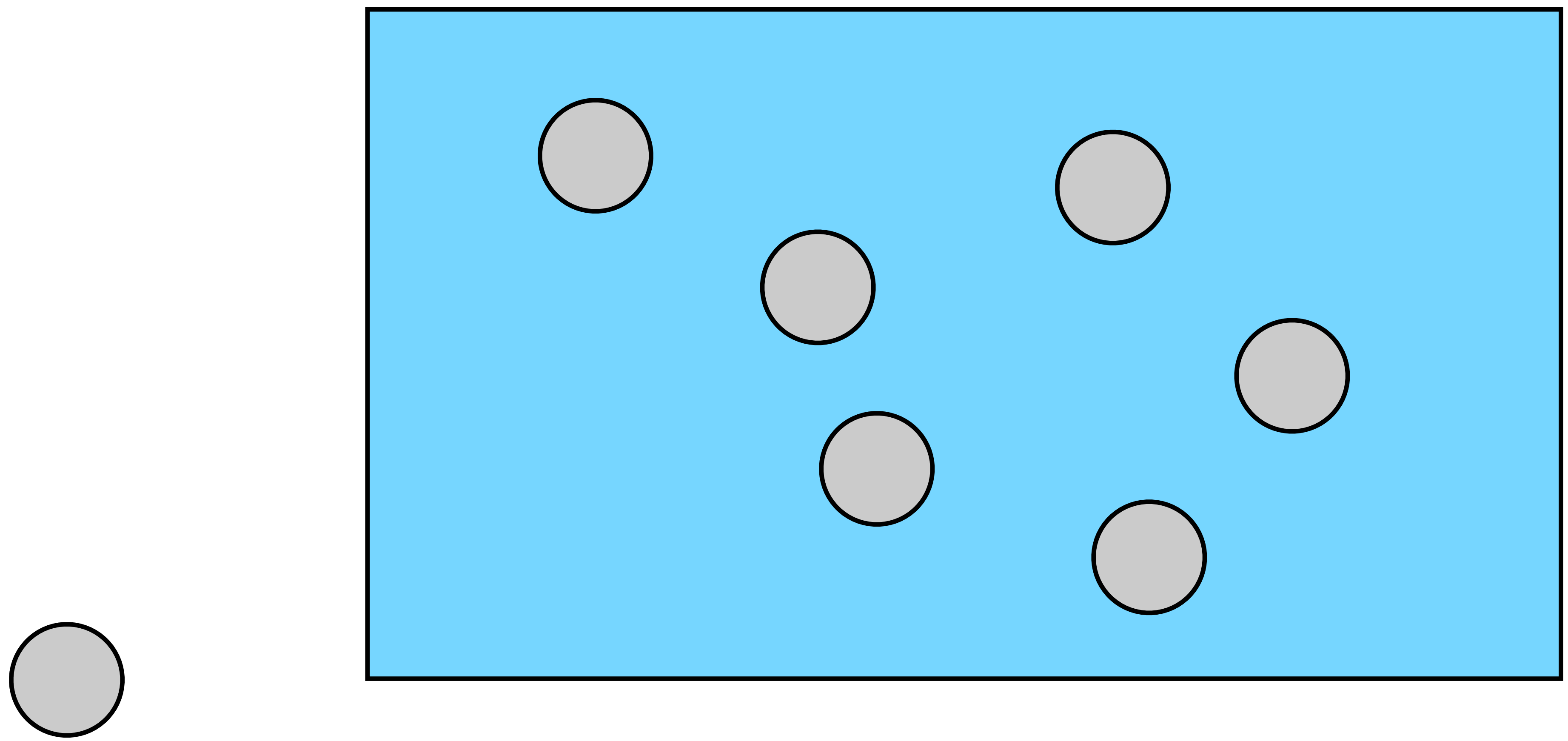
3. Adults speak

Population model



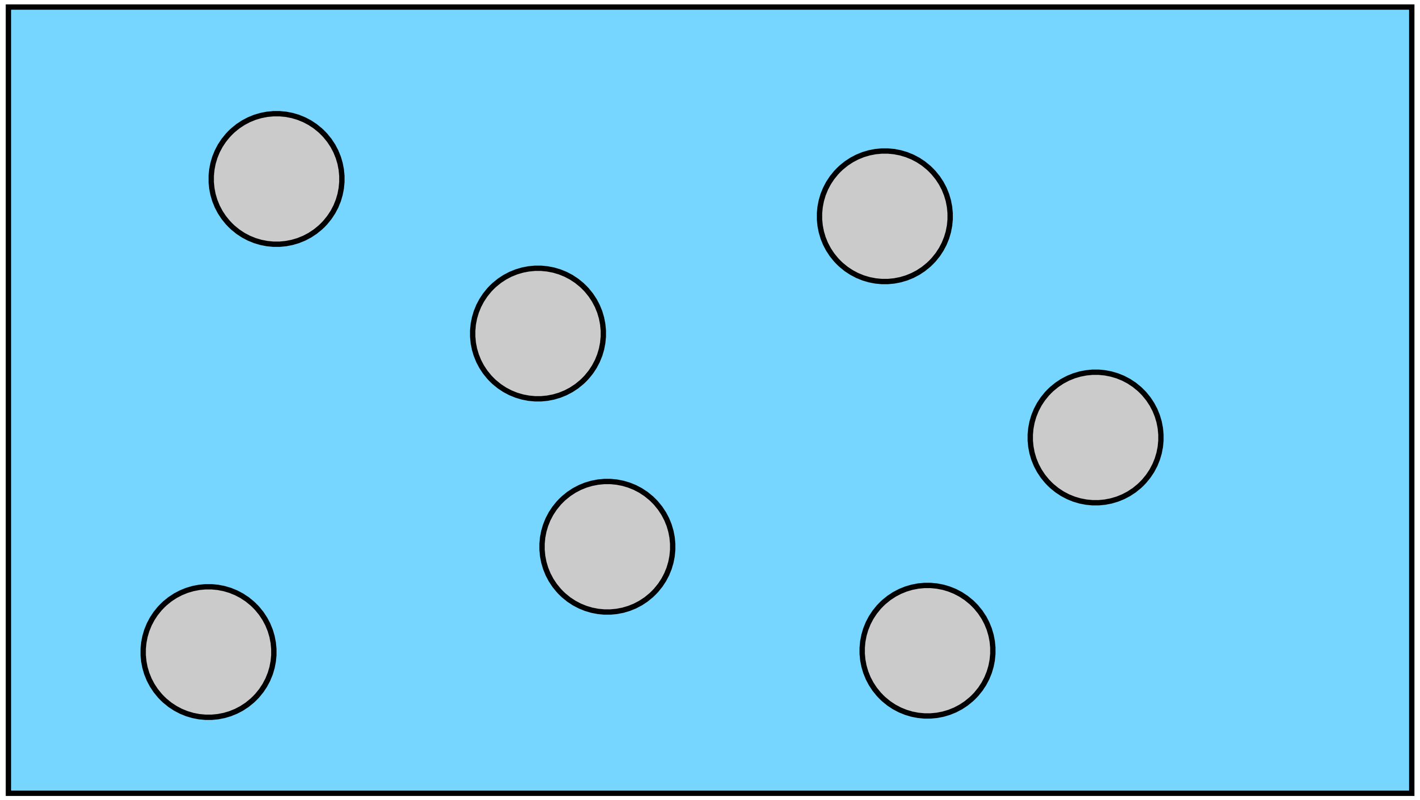
4. Child learns

Population model



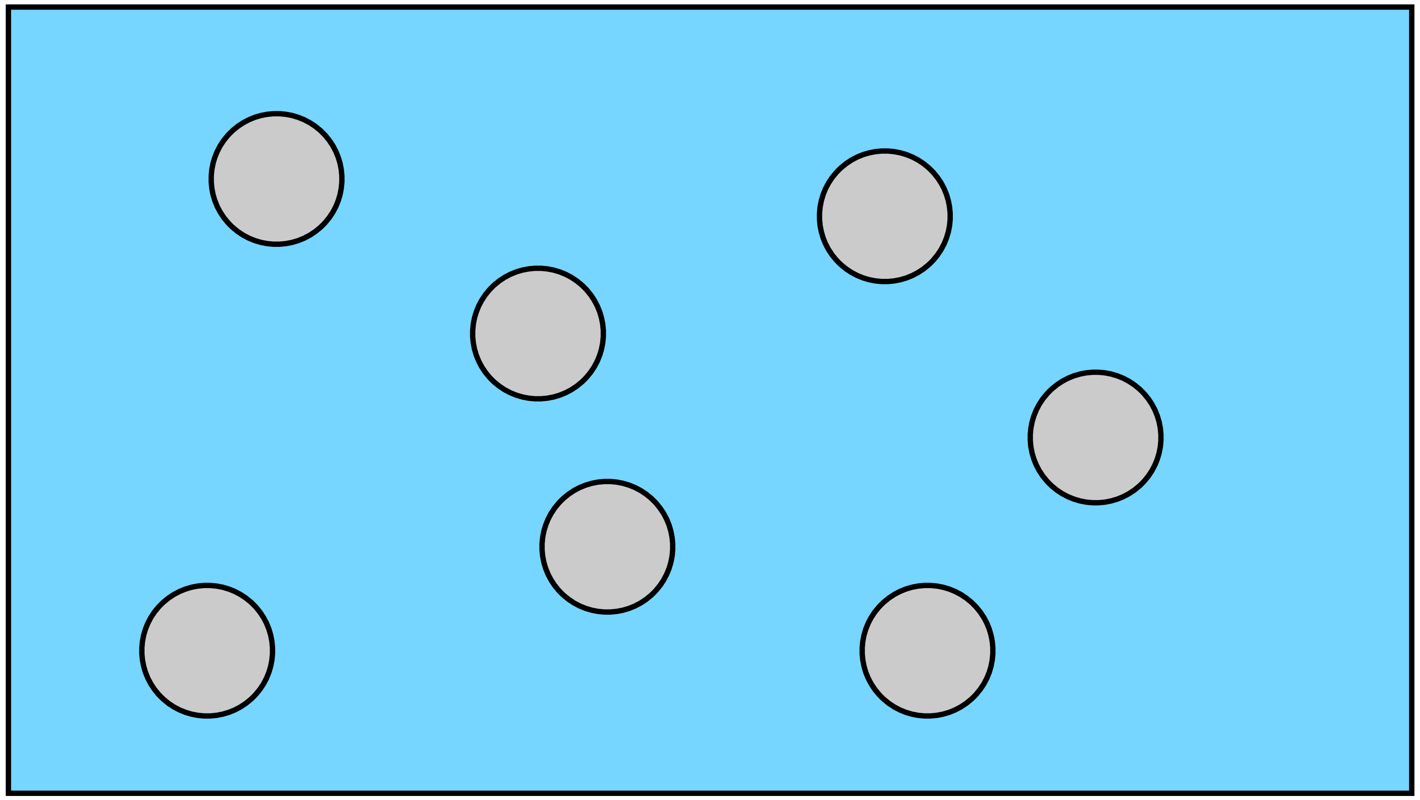
5. Child enters population

Population model



5. Child enters population

Population model



6. Repeat

An aside: other possible population models

An aside: other possible population models

- Mesoudi & Whiten (2008) review a number of different ways we could model cultural evolution.

An aside: other possible population models

- Mesoudi & Whiten (2008) review a number of different ways we could model cultural evolution.

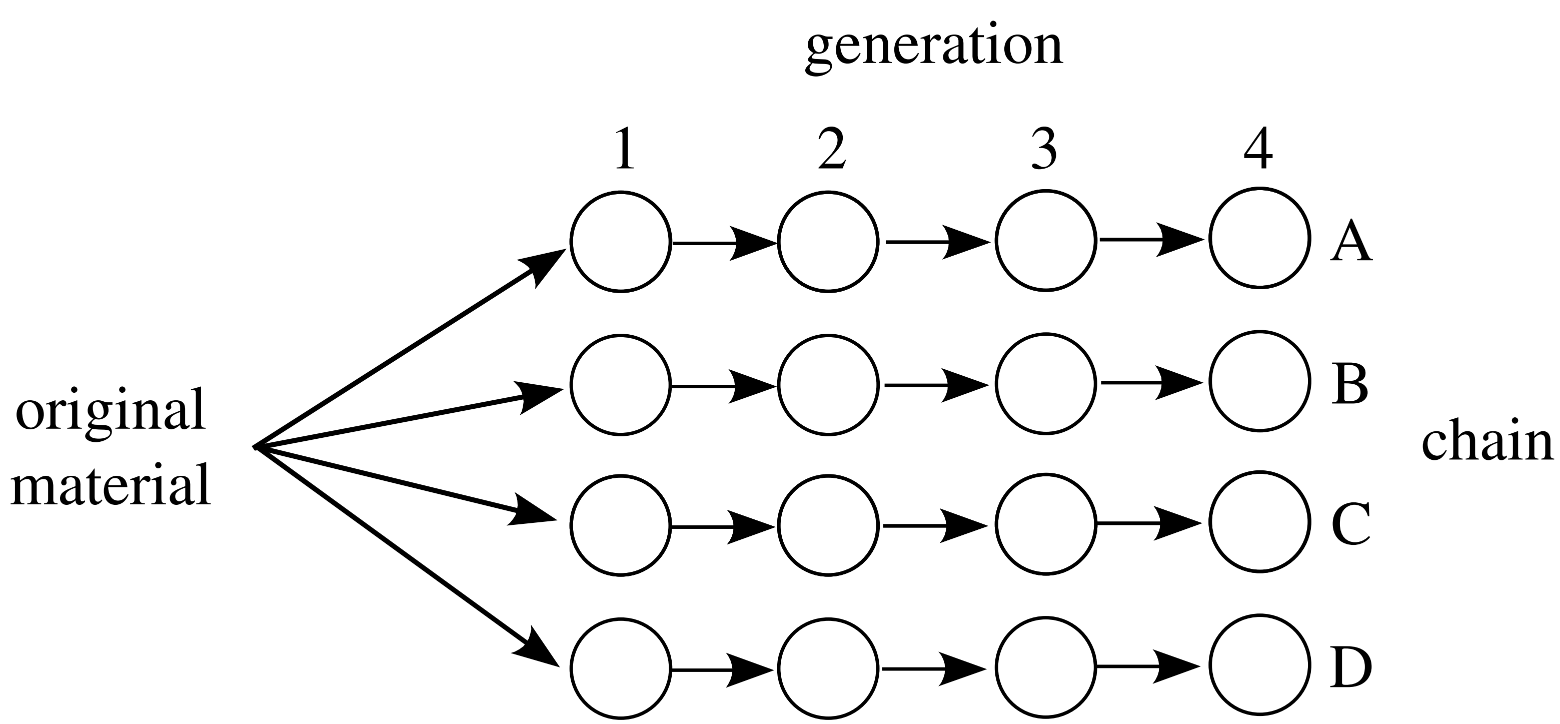


Figure 1. Design of a typical transmission chain study. The original material is passed along parallel chains of participants (represented by circles). Here, there are four chains (A–D), each comprising four generations (1–4). Adapted from Mesoudi (2007).

Transmission Chain

An aside: other possible population models

- Mesoudi & Whiten (2008) review a number of different ways we could model cultural evolution.

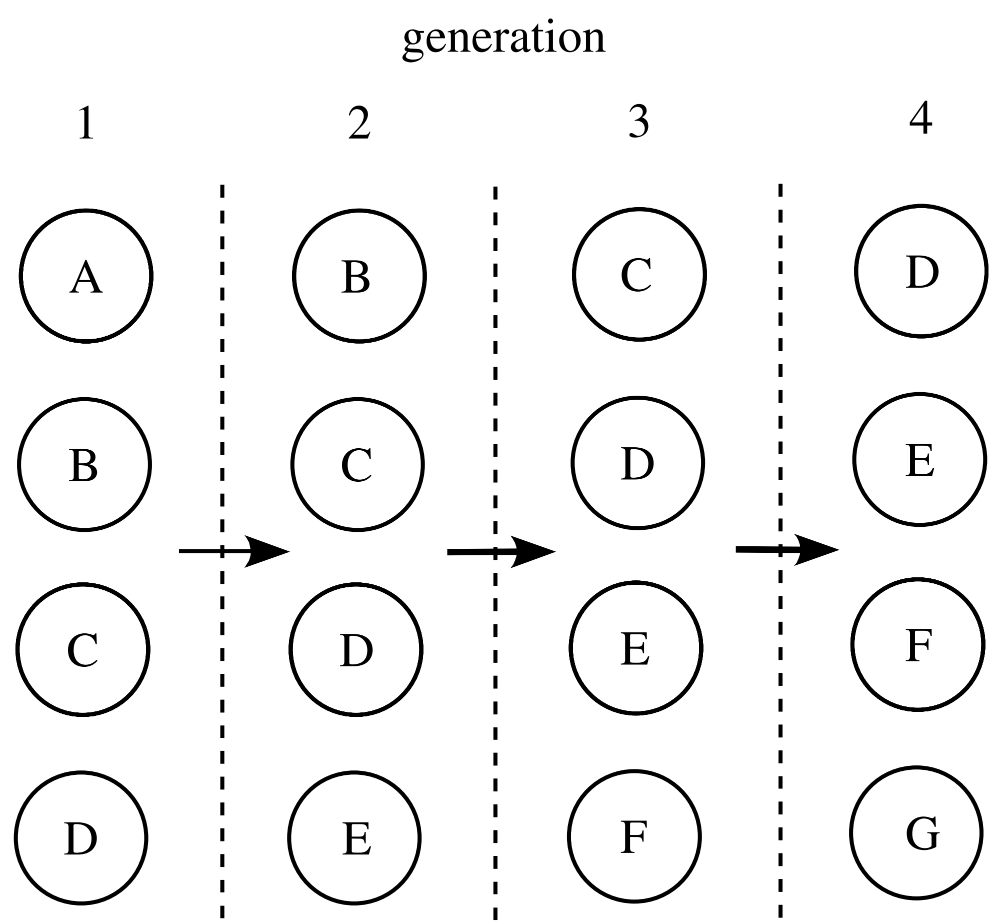


Figure 2. Design of a typical replacement study. Four participants (A–D) engage in a learning task, and in each generation one member of the group is replaced with a new participant. Adapted from Mesoudi (2007).

Replacement method

An aside: other possible population models

- Mesoudi & Whiten (2008) review a number of different ways we could model cultural evolution.

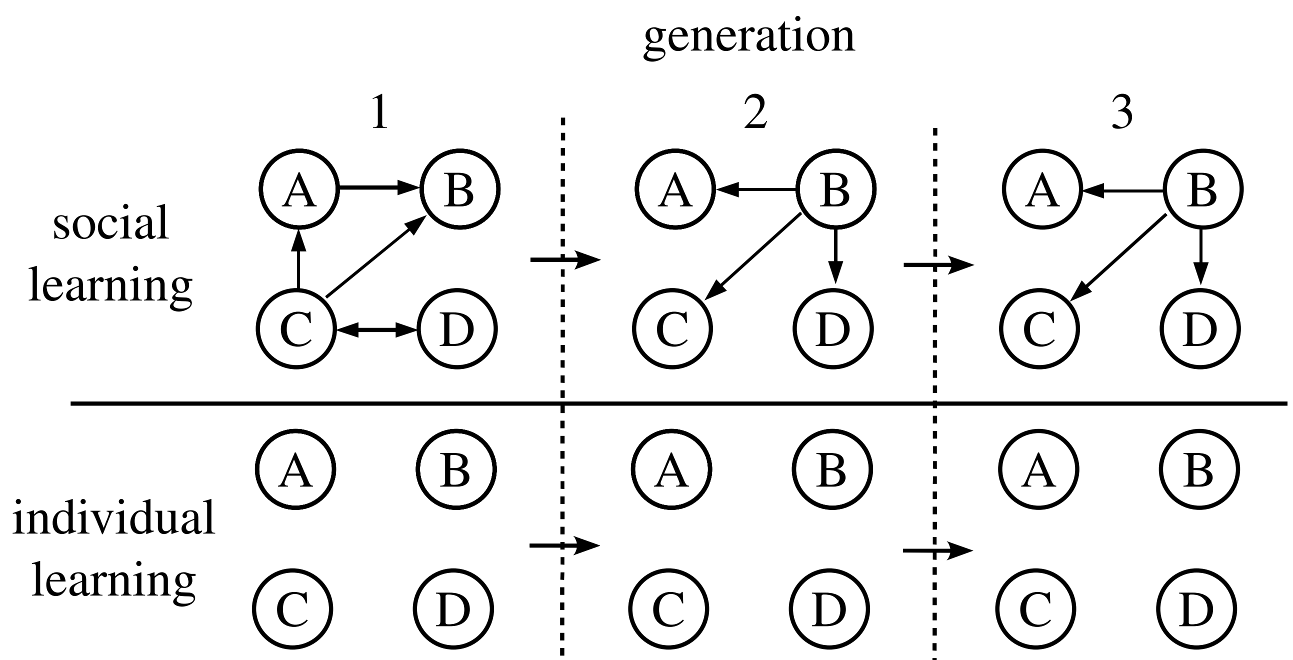


Figure 3. Design of a typical closed-group study. In the social learning condition, four participants (A–D) repeatedly engage in a learning task. Arrows indicate the flow of information via social learning, e.g. in generation 1, A learns from C, B learns from A and C, and C and D learn from each other. In generations 2 and 3, A, C and D all learn from B, who might have been recognized (or manipulated) to be particularly successful or prestigious. In the individual learning control condition, four participants engage in the same task but with no social interaction. Adapted from Mesoudi (2007).

Closed group

Beyond the acquisition test

Beyond the acquisition test

- Using *replacement method*, Smith (2002) set out two further tests: **maintenance** & **construction**

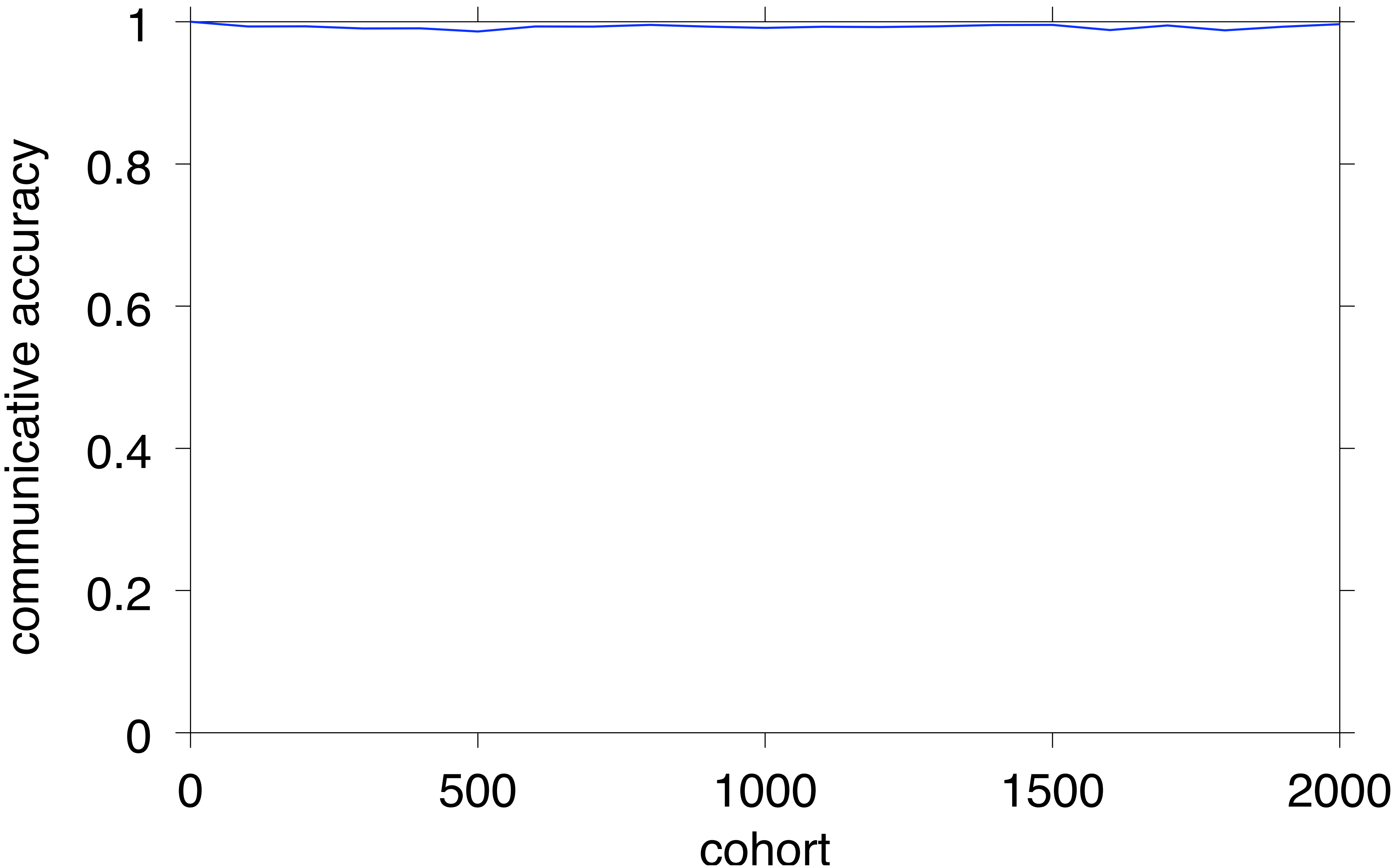
Beyond the acquisition test

- Using *replacement method*, Smith (2002) set out two further tests: **maintenance** & **construction**
- Both test behaviour over time *in a population* through iterated learning (i.e. cultural evolution of language)
- Differ only in their starting condition:
  - **maintenance** - starts with optimal system
  - **construction** - starts with random system

Maintenance

- Can a population using our basic rule [1, 0, 0, 0] maintain an optimal system? (under noisy conditions)
  - Note that Smith (2002) uses communicative accuracy as an evaluation metric for this
  - In other words, answer this question by testing whether the population can communicate successfully

Result: yes



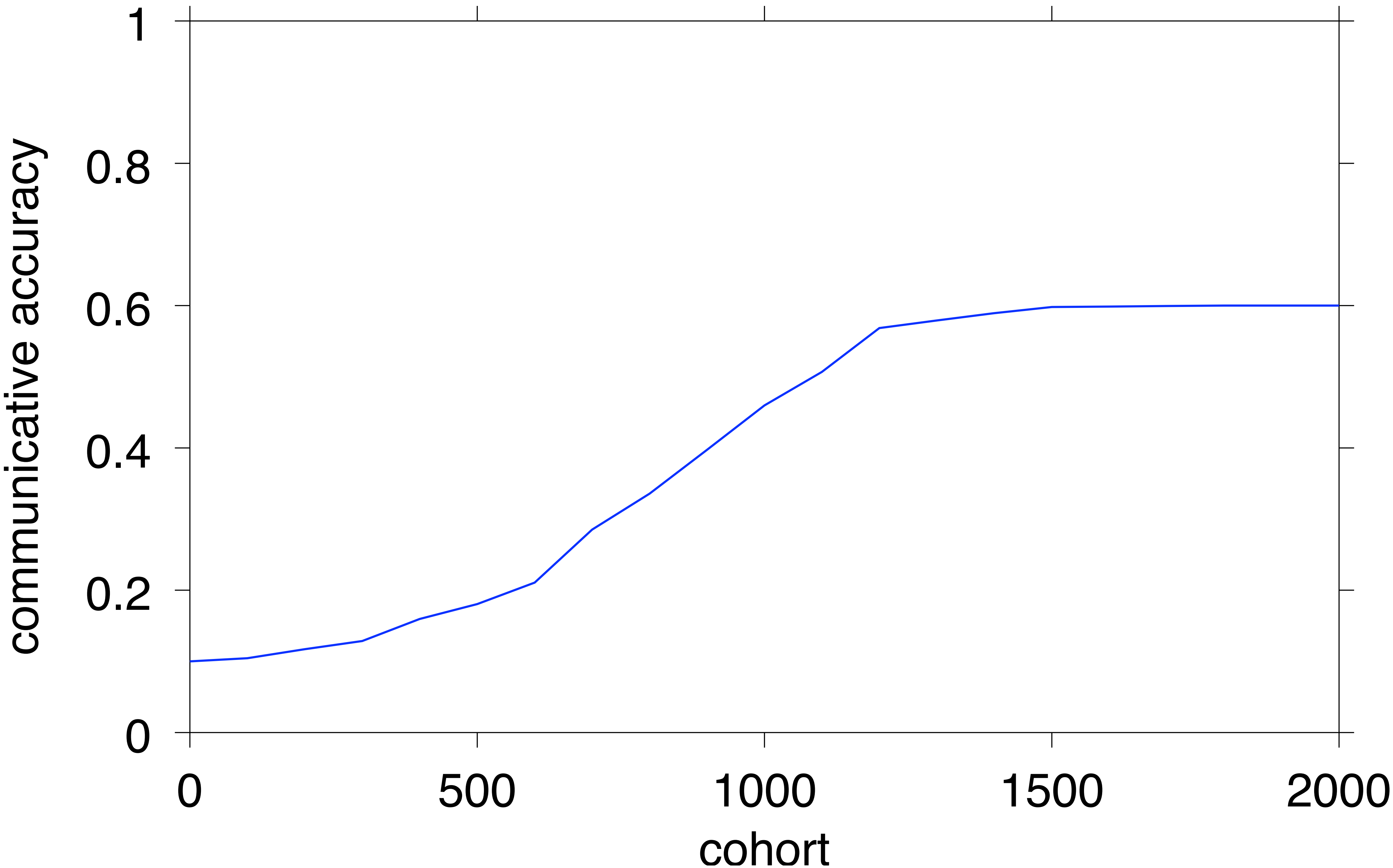
communicative accuracy
0
0.2
0.4
0.6
0.8
1
0
500
1000
1500
2000
cohort

Construction test

- Can a population using our rule [1,0,0,0] construct an optimal system from randomness?

- In other words, can our model not only account for learning, but also the cultural evolution of optimal communication in the first place?

Result: no


1
0.8
0.6
0.4
0.2
0
communicative accuracy
0
500
1000
1500
2000
cohort

What have we discovered?

- If individuals learn like this:

$\Delta w_{m_i s_j}$ = **+1** if both m_i and s_j are active
$\Delta w_{m_i s_j}$ = **0** if m_i is active and s_j is inactive
$\Delta w_{m_i s_j}$ = **0** if m_i is inactive and s_j is active
$\Delta w_{m_i s_j}$ = **0** if both m_i and s_j are inactive

- Individuals can **learn** an optimal system
- Populations will **maintain** an optimal system
- Populations *cannot* **construct** an optimal system

Another weight update rule

- How about this one?

$\Delta w_{m_i s_j}$ = **+1** if both m_i and s_j are active
$\Delta w_{m_i s_j}$ = **-1** if m_i is active and s_j is inactive
$\Delta w_{m_i s_j}$ = **-1** if m_i is inactive and s_j is active
$\Delta w_{m_i s_j}$ = **+1** if both m_i and s_j are inactive

- Is this fundamentally the same as the old rule?
  Different? Better? Worse?

Another weight update rule

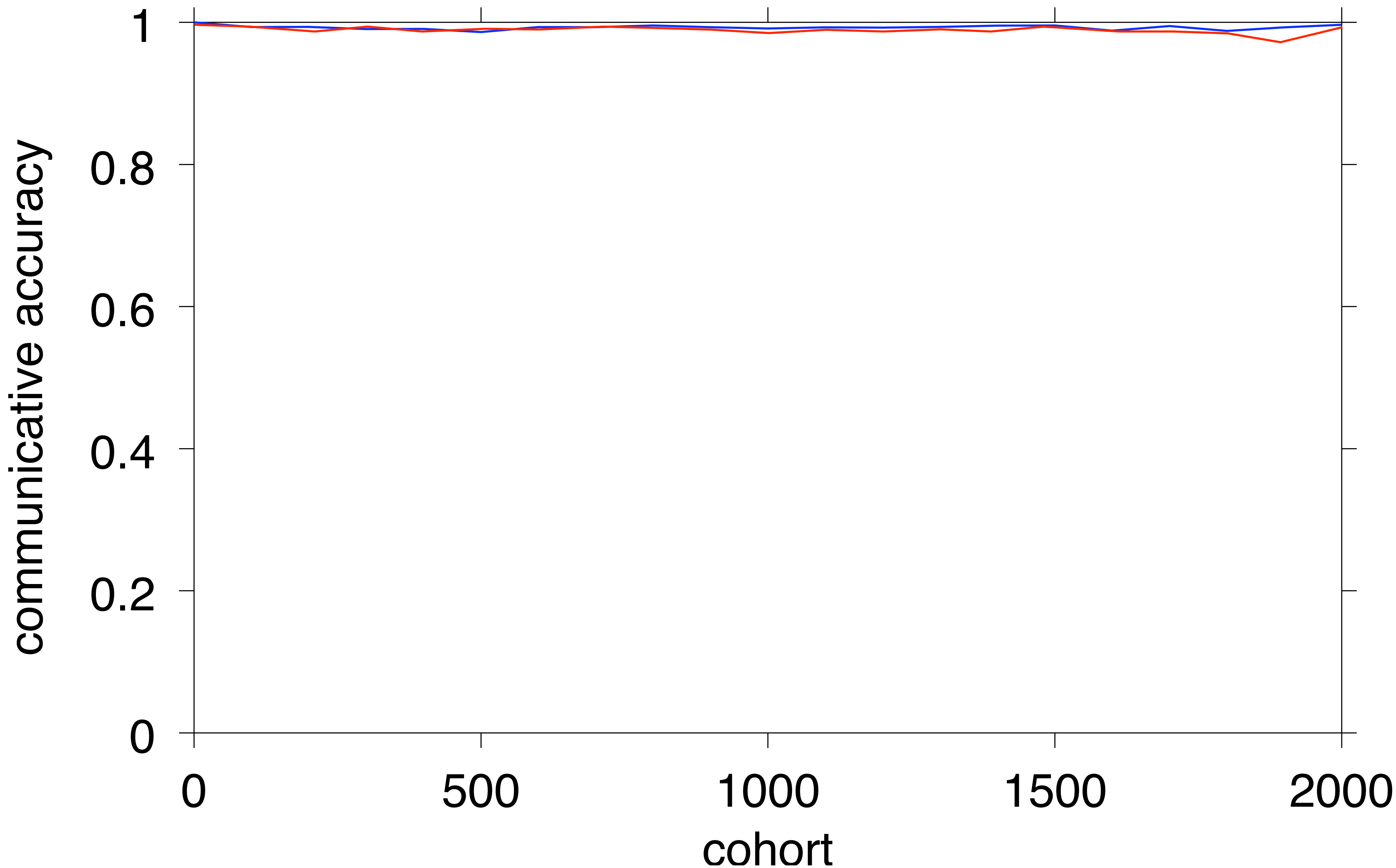
- How about this one?

$\Delta w_{m_i s_j}$ = **+1** if both m_i and s_j are active
$\Delta w_{m_i s_j}$ = **-1** if m_i is active and s_j is inactive
$\Delta w_{m_i s_j}$ = **-1** if m_i is inactive and s_j is active
$\Delta w_{m_i s_j}$ = **+1** if both m_i and s_j are inactive

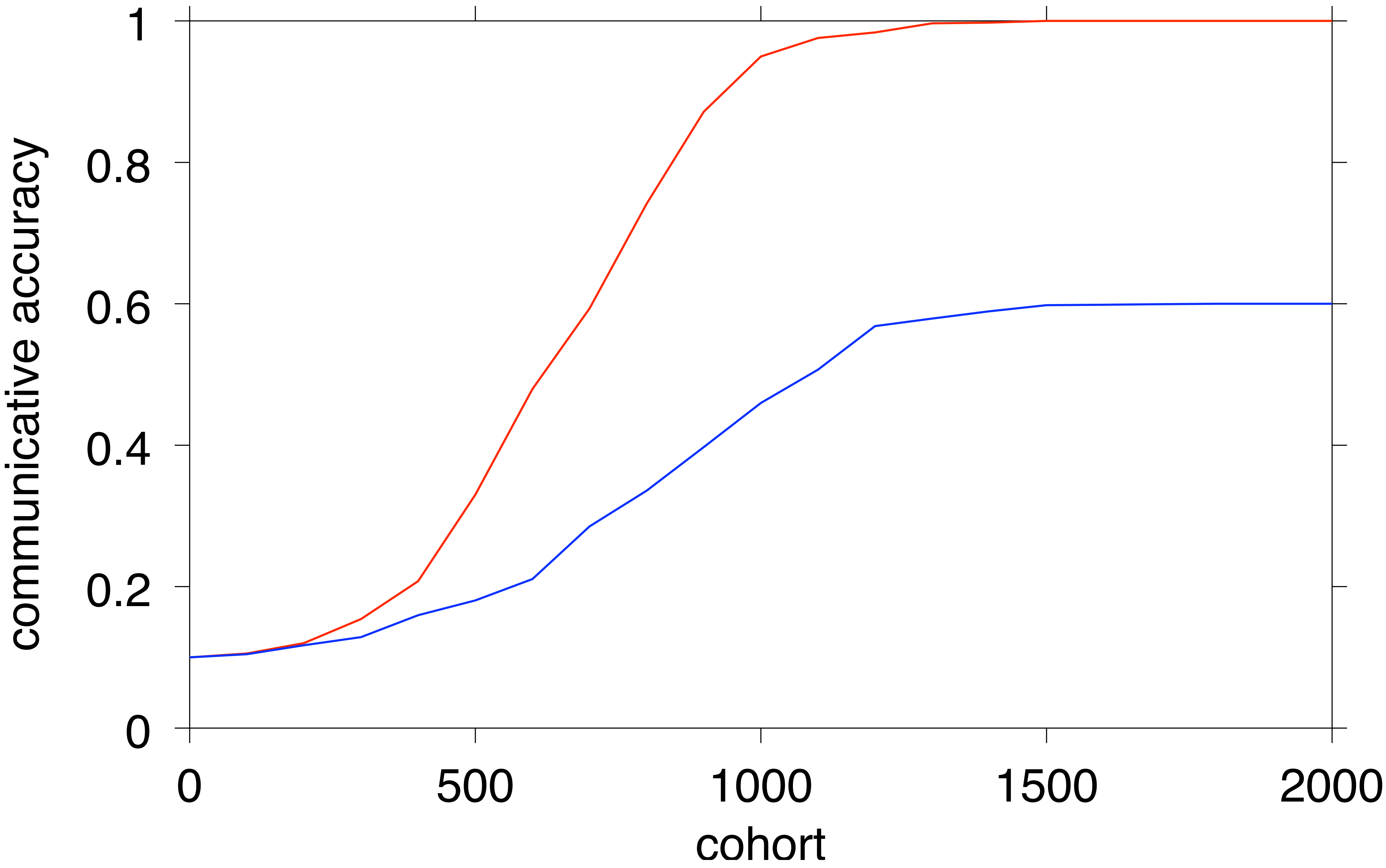
- Is this fundamentally the same as the old rule?
  Different? Better? Worse?

- **Acquisition test**: yes, an individual with this rule can learn an optimal system

Maintenance test: yes (again)



Construction: yes!



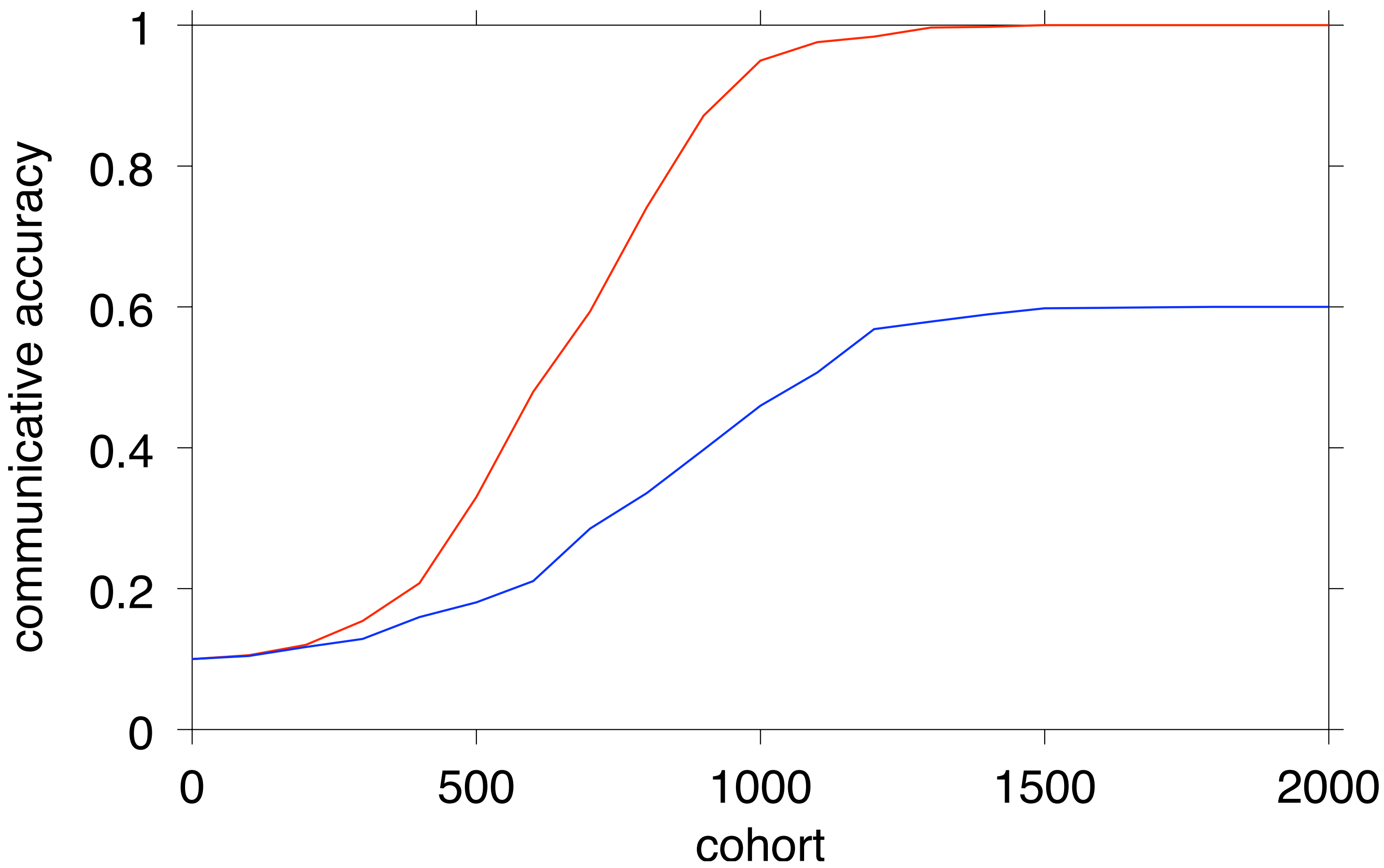
What have we discovered?

- If individuals learn like this:

$\Delta w_{m_i s_j}$ = **+1** if both m_i and s_j are active
$\Delta w_{m_i s_j}$ = **-1** if m_i is active and s_j is inactive
$\Delta w_{m_i s_j}$ = **-1** if m_i is inactive and s_j is active
$\Delta w_{m_i s_j}$ = **+1** if both m_i and s_j are inactive

- Individuals can **learn** an optimal system
- Populations will **maintain** an optimal system
- Populations will **construct** an optimal system

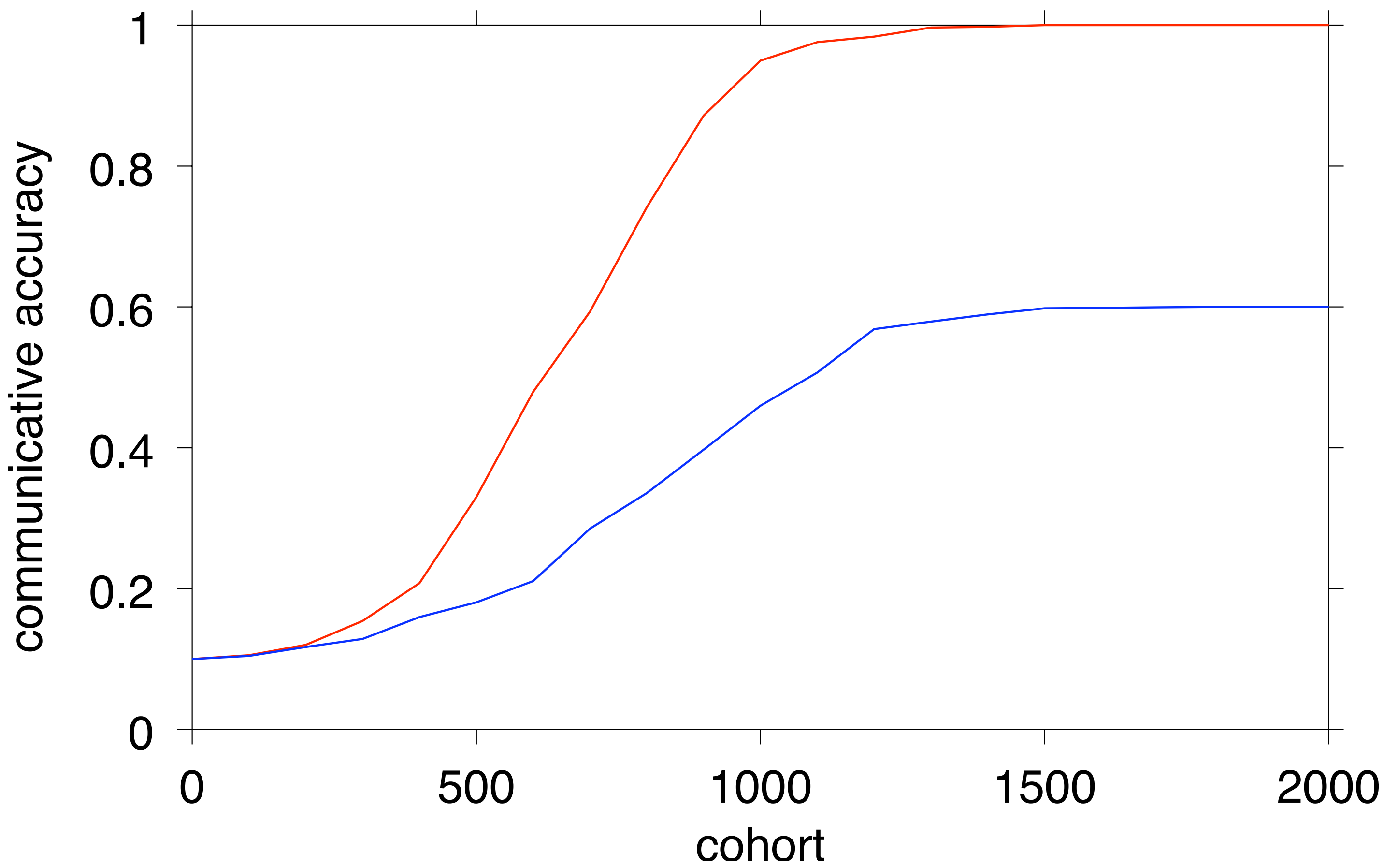
The difference



The only difference is the way they learn

- Language changes (evolves) **because of the way individuals learn**

The difference



The only difference is the way they learn

- Language changes (evolves) **because of the way individuals learn**
- Different ways of learning give different (cultural) evolutionary outcomes, even if the behaviour is the same at the individual level!

What about this rule?

$\Delta w_{m_i s_j}$ = **+1** if both m_i and s_j are active

$\Delta w_{m_i s_j}$ = **-1** if m_i is active and s_j is inactive

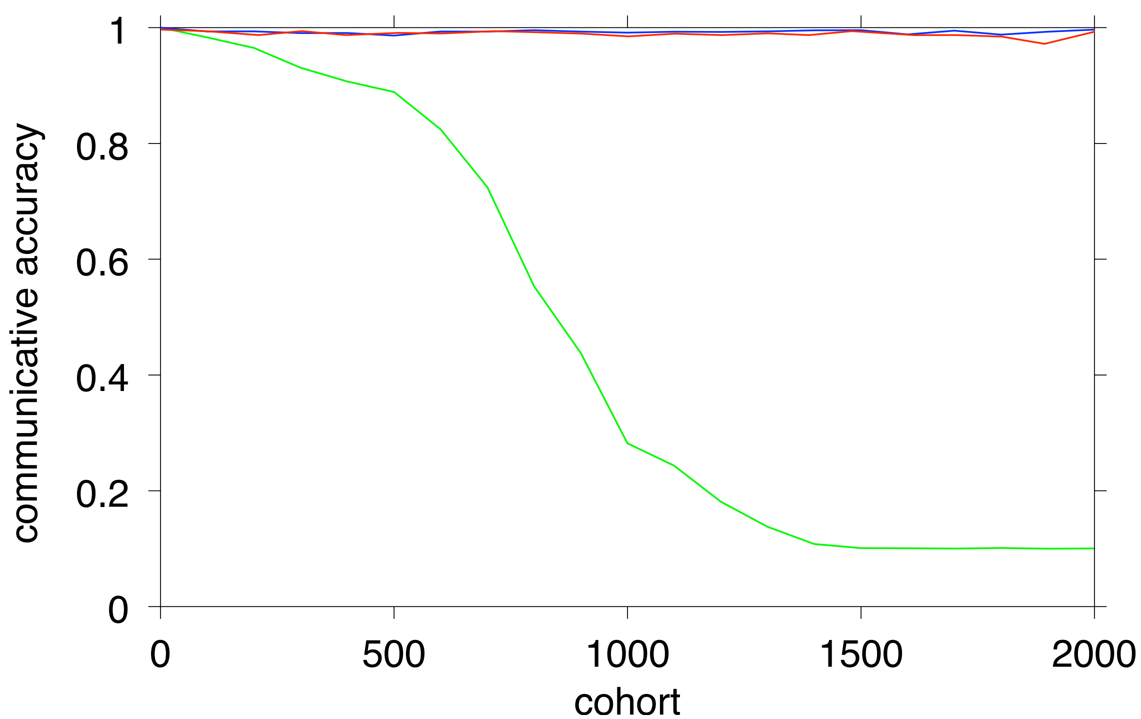
$\Delta w_{m_i s_j}$ = **+1** if m_i is inactive and s_j is active

$\Delta w_{m_i s_j}$ = **0** if both m_i and s_j are inactive

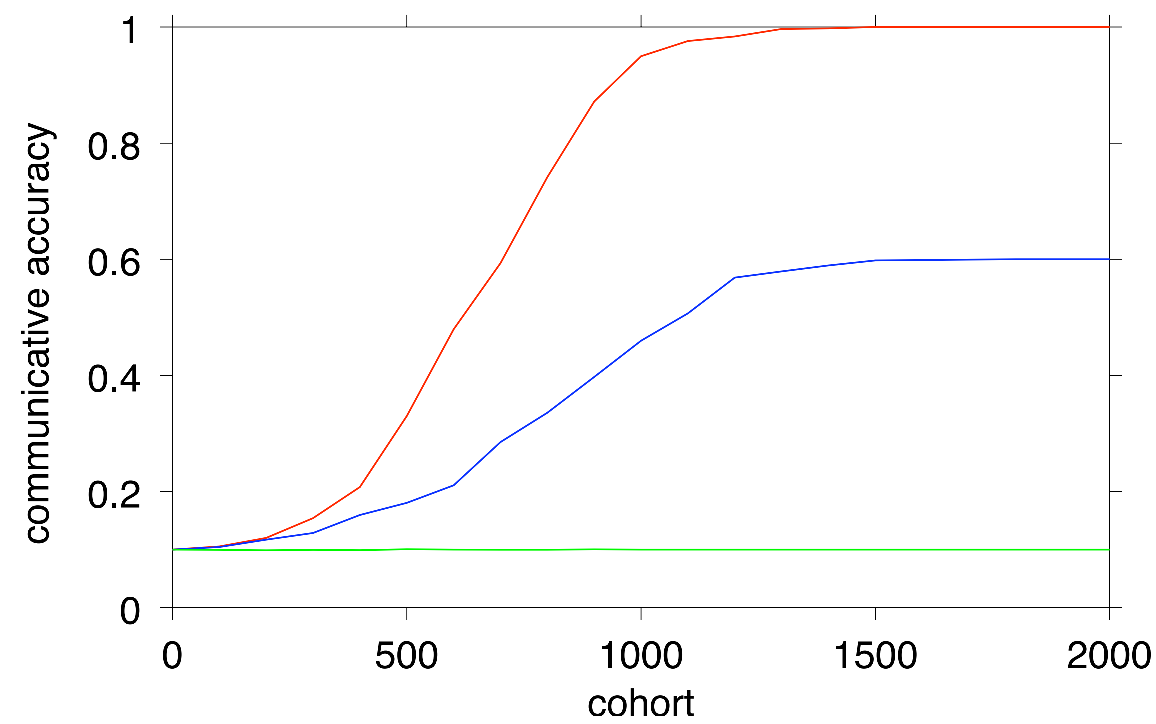
Results for third rule: [1, -1, 1, 0]

Passes acquisition test? **Yes**

Maintenance: **No**



Construction: **No**



What about all 81 possible rules?

What about all 81 possible rules?

- For all the values of α, β, γ and δ, there may be different learning biases

What about all 81 possible rules?

- For all the values of α, β, γ and δ, there may be different learning biases
- These may lead to different cultural evolutionary trajectories

What about all 81 possible rules?

- For all the values of α, β, γ and δ, there may be different learning biases
- These may lead to different cultural evolutionary trajectories
- Which ones allow for the evolution of language?

What about all 81 possible rules?

- For all the values of α, β, γ and δ, there may be different learning biases
- These may lead to different cultural evolutionary trajectories
- Which ones allow for the evolution of language?
- Which of these is closest to the real human learning bias?

What about all 81 possible rules?

- For all the values of α, β, γ and δ, there may be different learning biases
- These may lead to different cultural evolutionary trajectories
- Which ones allow for the evolution of language?
- Which of these is closest to the real human learning bias?
- What does this mean for how we think of the relationship between innateness, learning, and the explanation of the universals of language structure?

Readings for this class

- Smith, K. (2002). The cultural evolution of communication in a population of neural networks. *Connection Science,* 14, 65-84

- Mesoudi, A. & Whiten, A. (2008) The multiple roles of cultural transmission experiments in understanding human cultural evolution. *Phil Trans R. Soc. B*, 363, 3489-3501

    - Note that this deals with experiments with real human subjects rather than simulations, but the principles regarding population structures are just as valid.