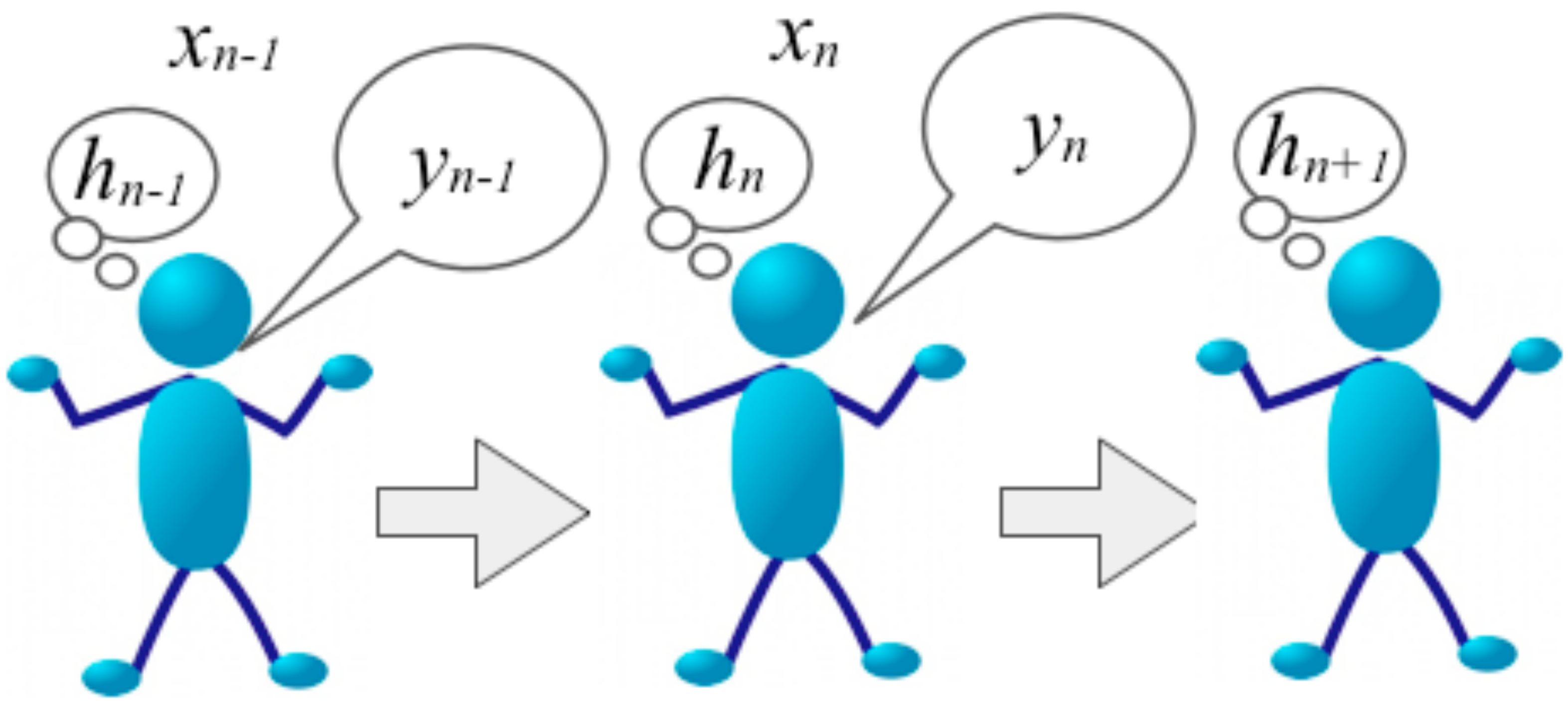
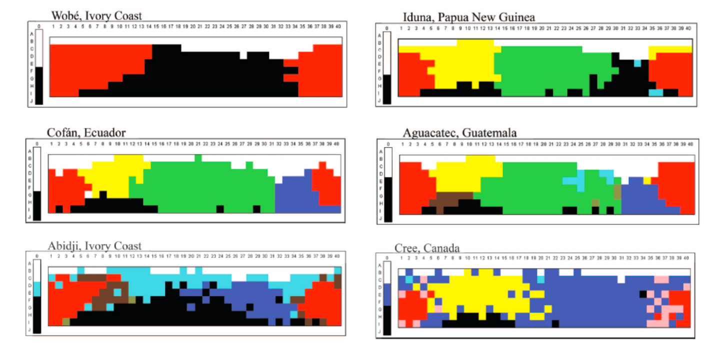
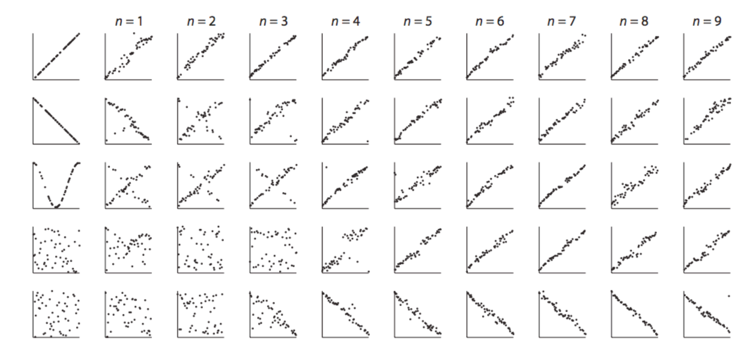
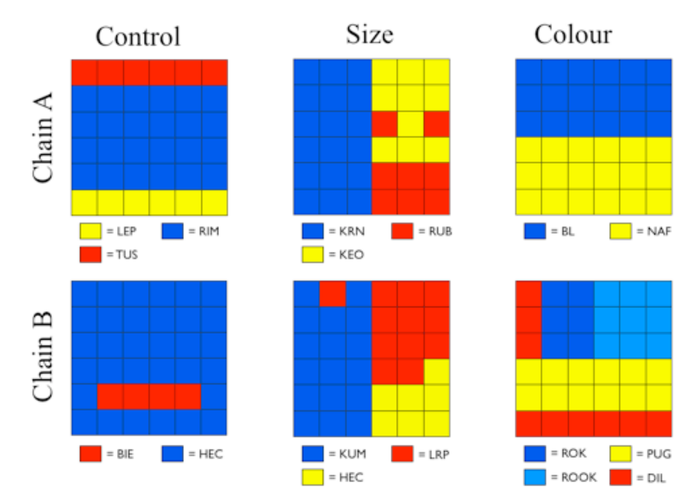
Computational Cognitive Science



Lecture 14: Iterated learning

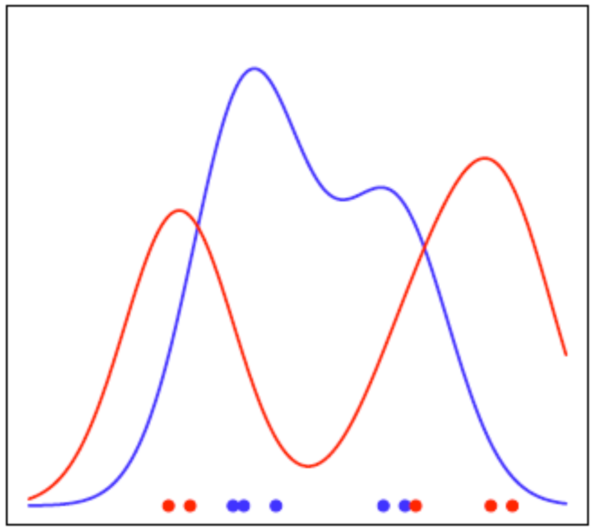


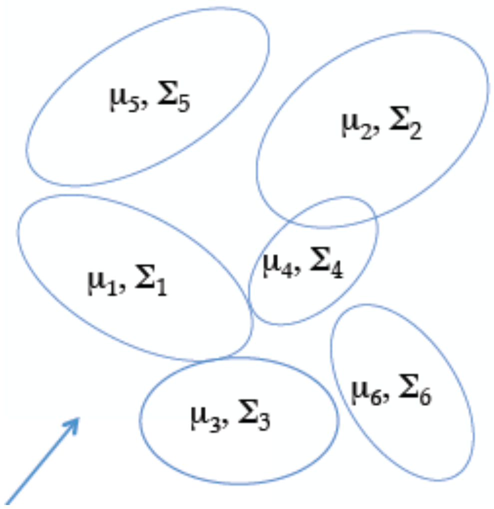


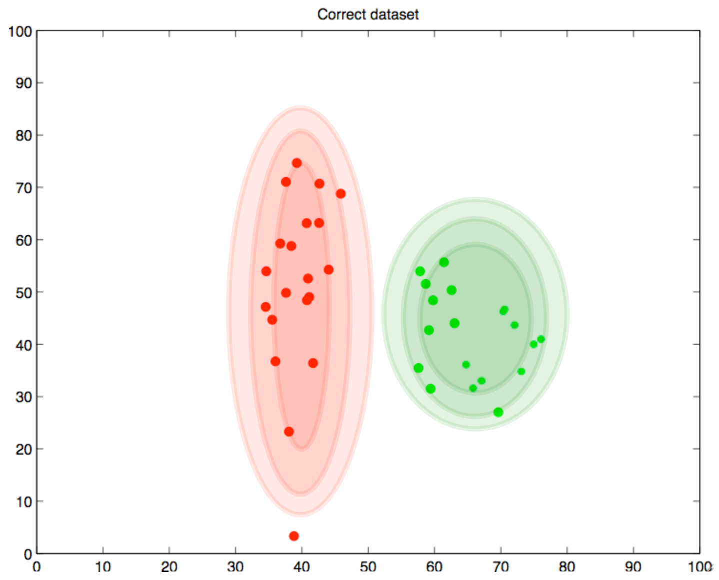


Let’s step back a bit

- So far we’ve spent all of our time exploring how people (or models) can learn concepts that don’t change, or incorporate an element of time
- Those concepts / models may be simple …

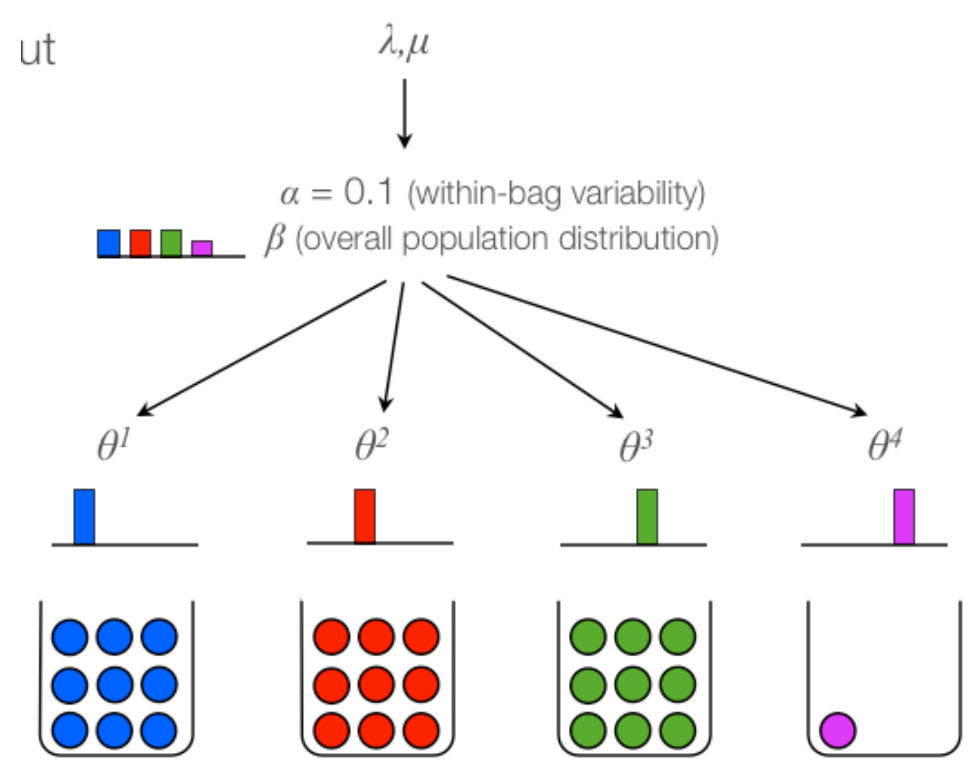






Let's step back a bit

- So far we've spent all of our time exploring how people (or models) can learn concepts that don't change, or incorporate an element of time
- Those concepts / models may be simple ...
- or more complicated ....



Let’s step back a bit

- So far we’ve spent all of our time exploring how people (or models) can learn concepts that don’t change, or incorporate an element of time
- Those concepts / models may be simple ...
- or more complicated ....
- But what about structure and knowledge that occurs over time?

Let’s step back a bit

- So far we’ve spent all of our time exploring how people (or models) can learn concepts that don’t change, or incorporate an element of time
- Those concepts / models may be simple ...
- or more complicated ....
- But what about structure and knowledge that occurs over time?

- Today: How did concepts and ideas themselves evolve over time to be the way they are?

Today's plan

- Evidence for conceptual evolution
  - Inevitable given noisy transmission
  - Historical record
  - Cultural variation
- A model of conceptual change over time
  - Iterated learning model: basic idea
  - Mathematical proof and corresponding intuition
- Experimental evidence for iterated learning models
  - Function learning
  - Language
- Limitations and extensions to the iterated learning model
  - changing learner
  - changing producer
  - changing how hypotheses map onto the world

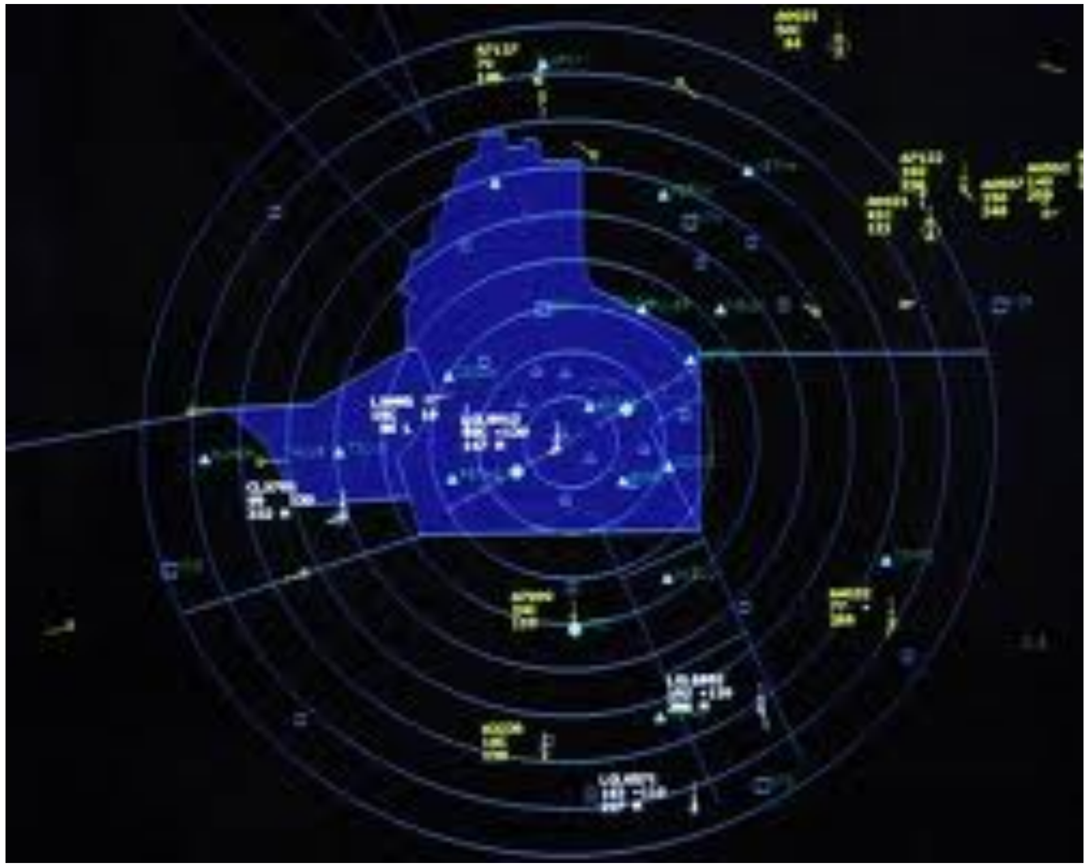
Today's plan

- ➡ Evidence for conceptual evolution
  - Inevitable given noisy transmission
  - Historical record
  - Cultural variation
- A model of conceptual change over time
  - Iterated learning model: basic idea
  - Mathematical proof and corresponding intuition
- Experimental evidence for iterated learning models
  - Function learning
  - Language
- Limitations and extensions to the iterated learning model
  - changing learner
  - changing producer
  - changing how hypotheses map onto the world

Why do we think our concepts evolved?

- Inevitable given noisy transmission
  - Anytime something must go through a process of *transmission* (which all concepts we learn from others do) that process affects the final outcome

aircraft controller -- sees lots of information,
needs to transmit it to pilots

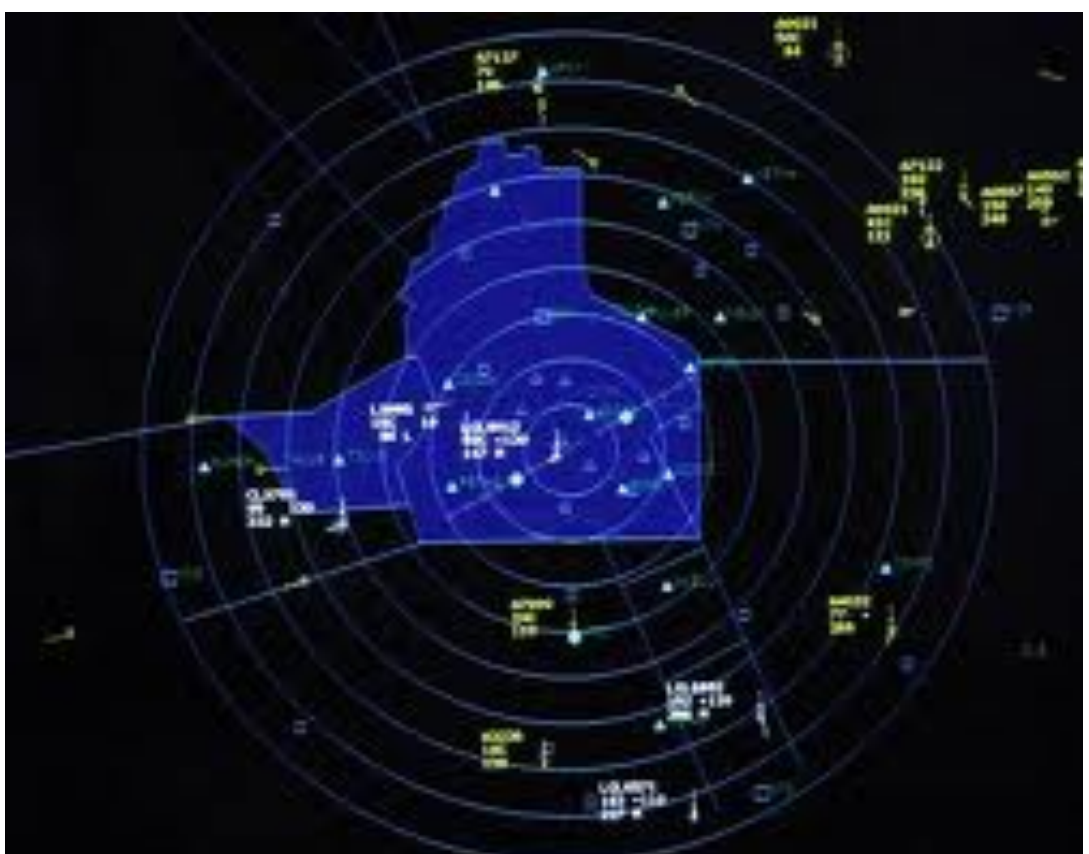


"Los Angeles tower, Mooney Niner One Seven Victor, cleared to land runway two five left"

Why do we think our concepts evolved?

- Inevitable given noisy transmission
  - Two sources of distortion: (a) introduction of noise into the transmission

aircraft controller -- sees lots of information,
needs to transmit it to pilots

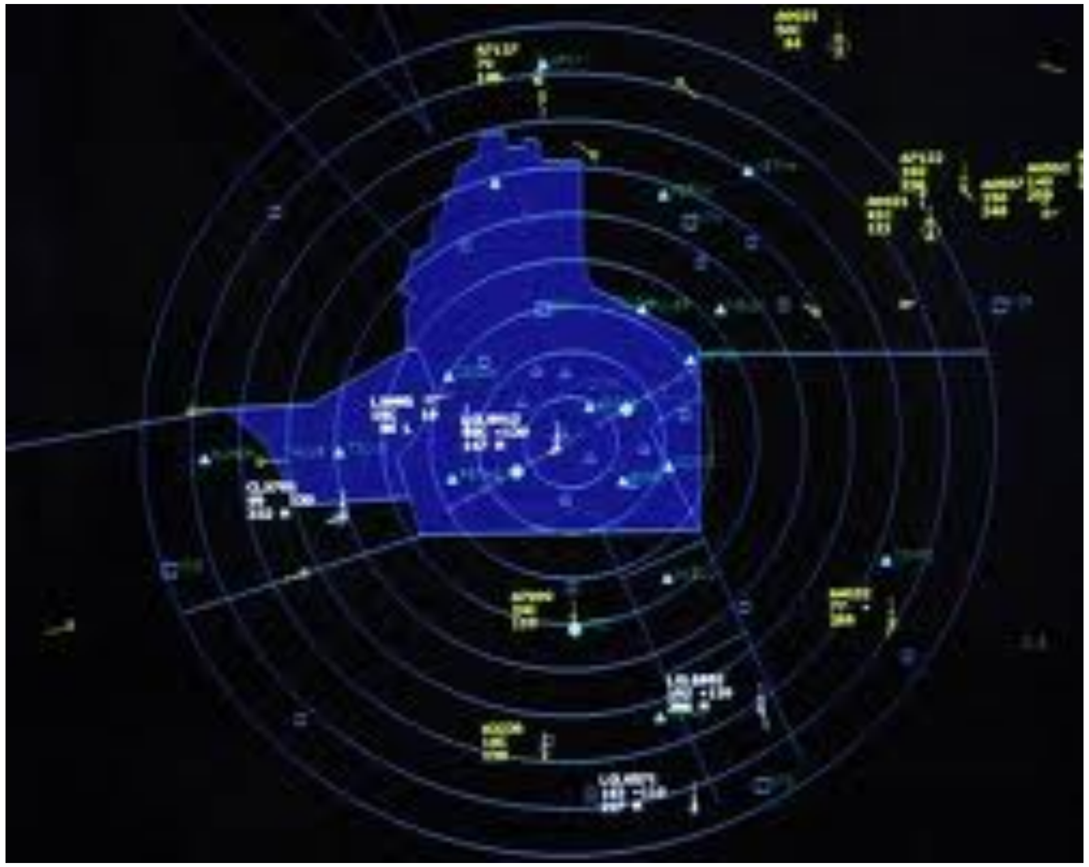


"Los Angeles tower,
Mooney N*** One Seven
Victor, cleared to ***
runway two five left"

Why do we think our concepts evolved?

- Inevitable given noisy transmission
  - Two sources of distortion: (a) introduction of noise into the transmission; (b) bottleneck on the amount of data you can transmit

aircraft controller -- sees lots of information, needs to transmit it to pilots



Okay, this is the controller from the Los Angeles tower calling. Flight Mooney Niner One Seven Victor, that is M917V, you're cleared to land on runway 25. That's the runway on the left, in between 24 and 26.

Why do we think our concepts evolved?

- Inevitable given noisy transmission
- Historical record
  - Lots of precedent for concepts changing over time

Why do we think our concepts evolved?

- Inevitable given noisy transmission
- Historical record
  - Lots of precedent for concepts changing over time

Ancient Greece/Rome:
- marriages arranged
- affairs (for men) okay, including with young boys
- not usually for love

medieval times:
- still usually economic
- often involved dowries
- women were property
- church involved more

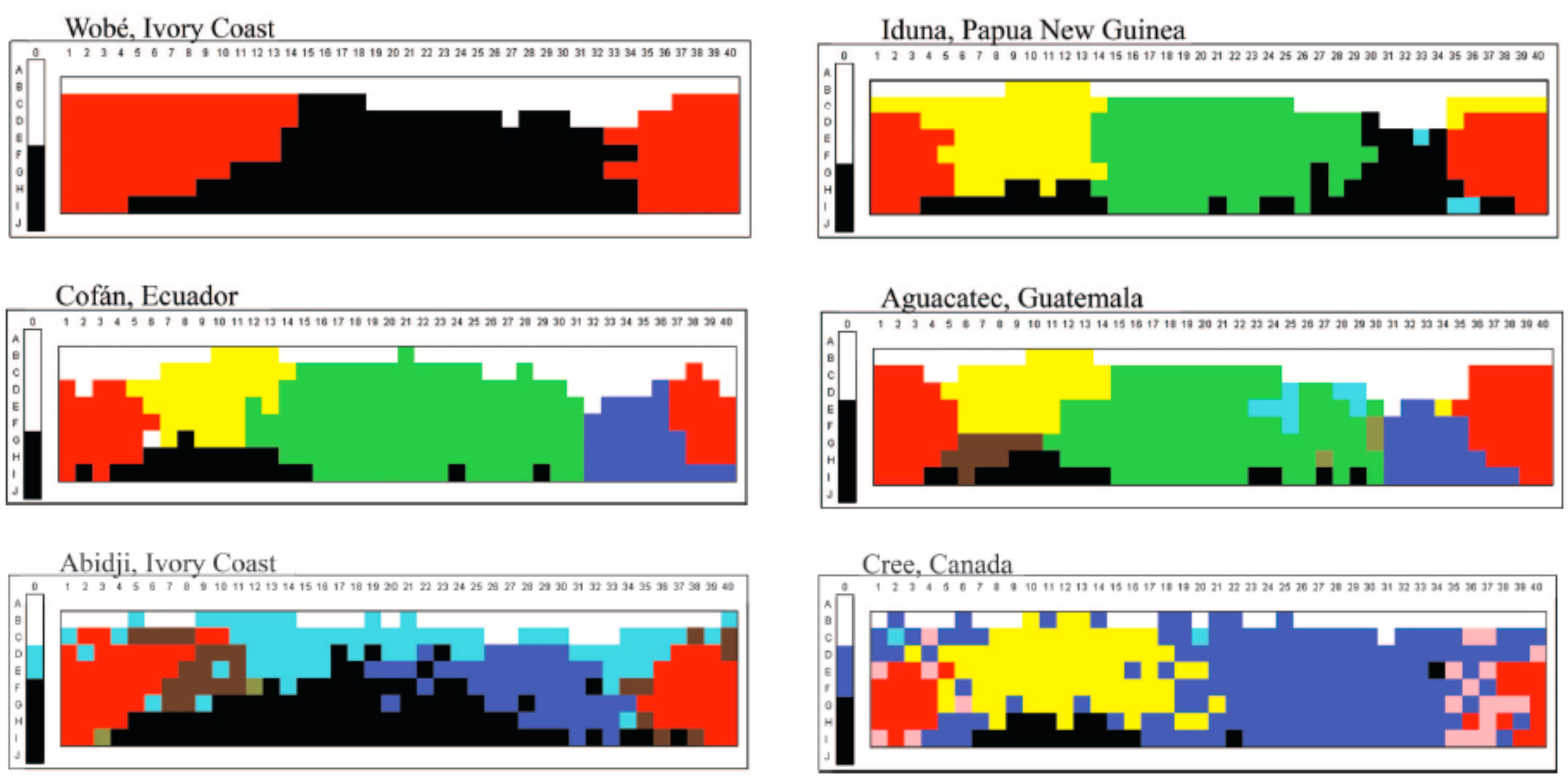
19th century times:
- occasionally for love
- often not cross-racial
- women sometimes can keep property

1950s etc
- often for love
- cross racial sometimes ok
- women “in the home”

today:
- usually for love
- cross-racial okay
- same-sex sometimes ok
- women wield much more economic power

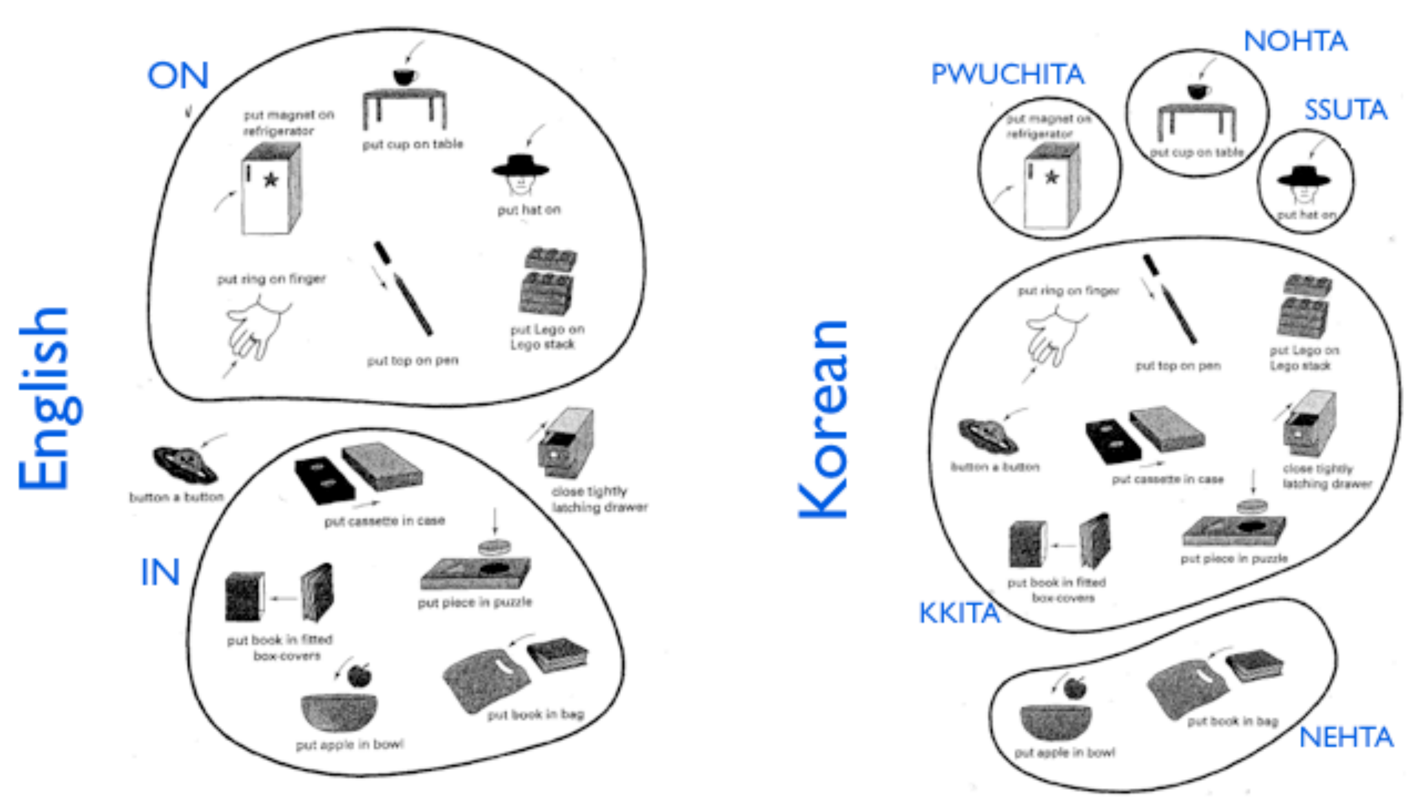
Why do we think our concepts evolved?

- Inevitable given noisy transmission
- Historical record
- Cultural variation
  - Many concepts vary between cultures



Why do we think our concepts evolved?

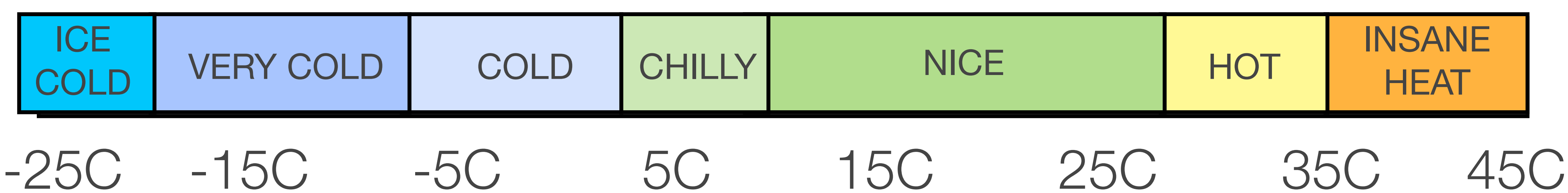
- Inevitable given noisy transmission
- Historical record
- Cultural variation
  - Many concepts vary between cultures



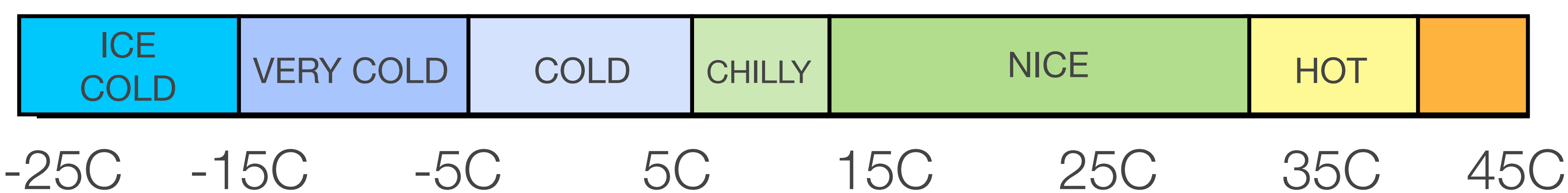
Why do we think our concepts evolved?

- Inevitable given noisy transmission
- Historical record
- Cultural variation
  - Many concepts vary between cultures

Colorado



Australia

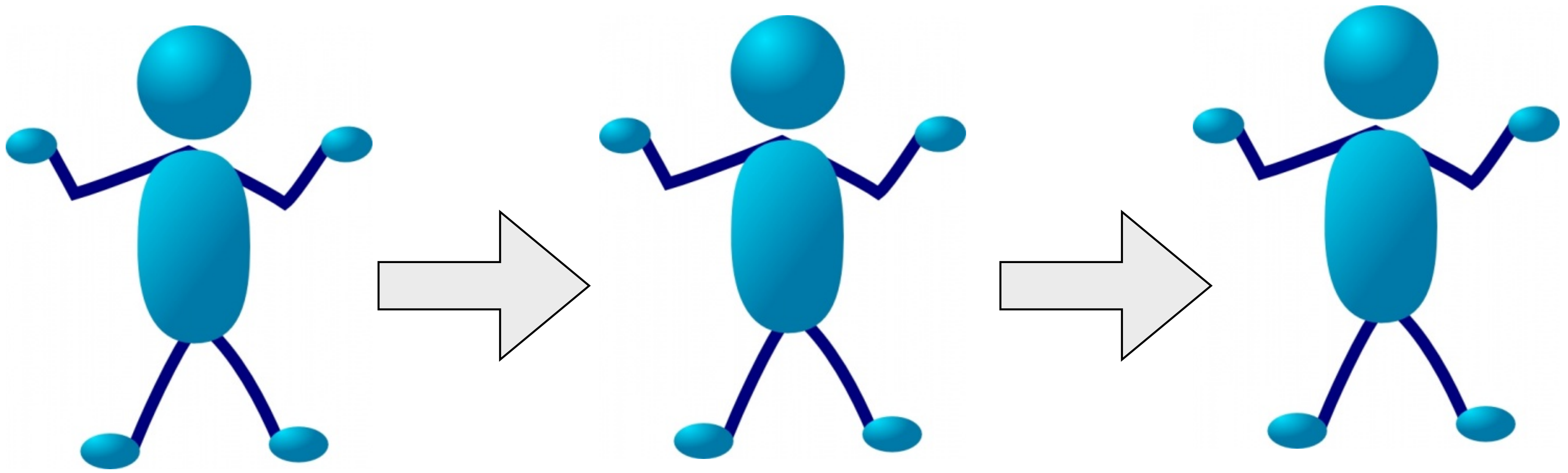


Today's plan

- Evidence for conceptual evolution
  - Inevitable given noisy transmission
  - Historical record
  - Cultural variation
- ➡ A model of conceptual change over time
  - Iterated learning model: basic idea
  - Mathematical proof and corresponding intuition
- Experimental evidence for iterated learning models
  - Function learning
  - Language
- Limitations and extensions to the iterated learning model
  - changing learner
  - changing producer
  - changing how hypotheses map onto the world

Iterated learning model: basic idea

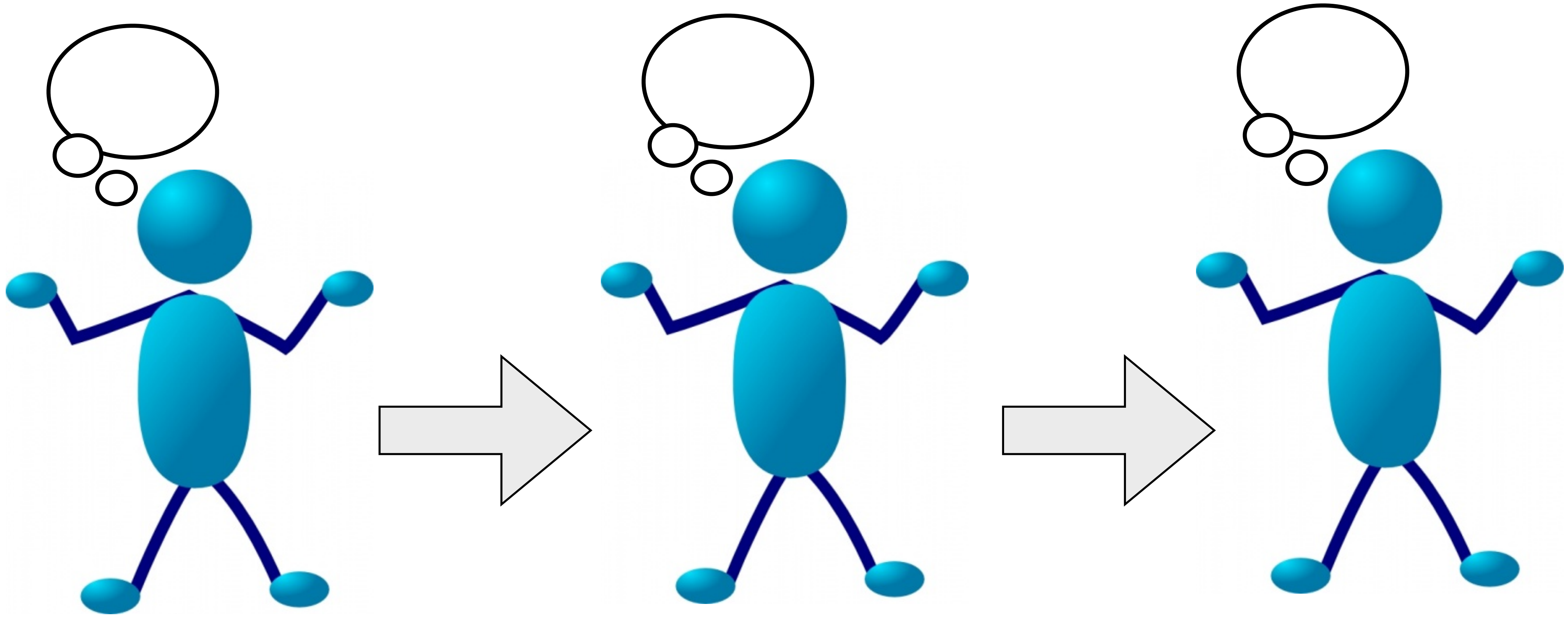
- Conceptual change in a population over time occurs through a process of *transmission*



Iterated learning model: basic idea

- Two main processes occur during this time that can shape how concepts change

cognition: how people learn from the data they see

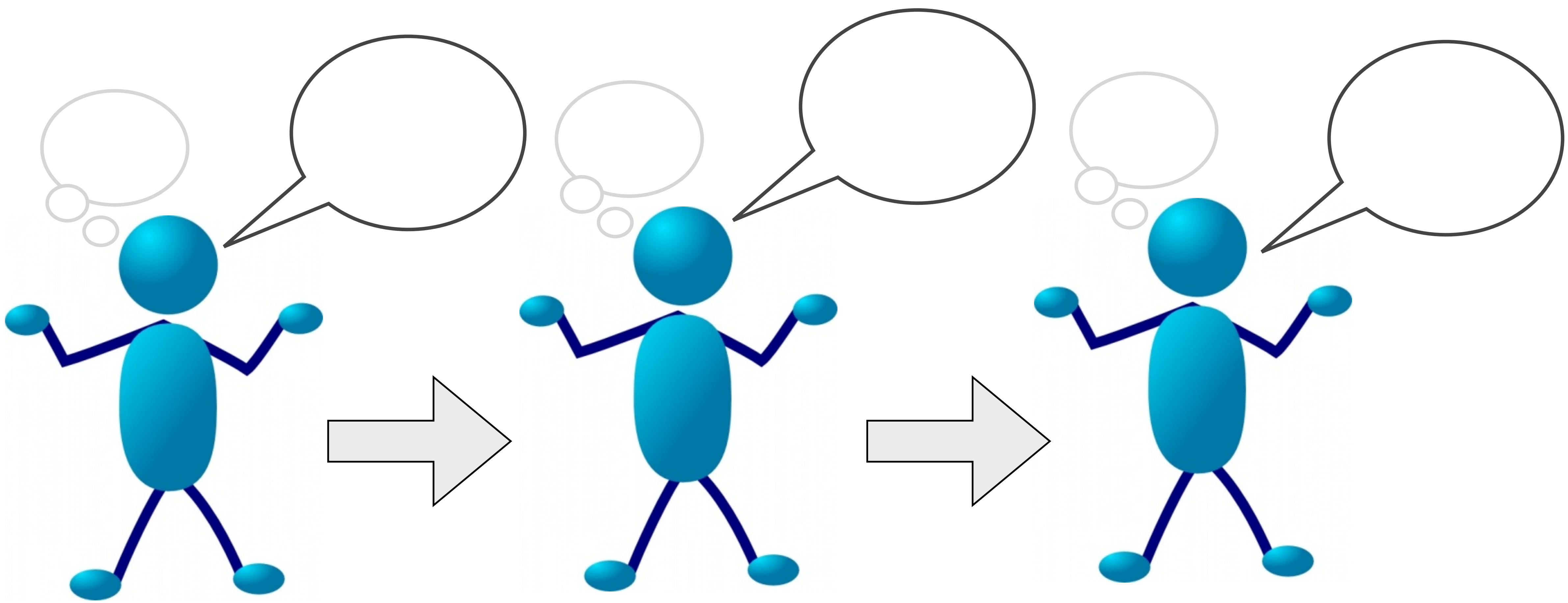


Iterated learning model: basic idea

- Two main processes occur during this time that can shape how concepts change

cognition: how people learn from the data they see

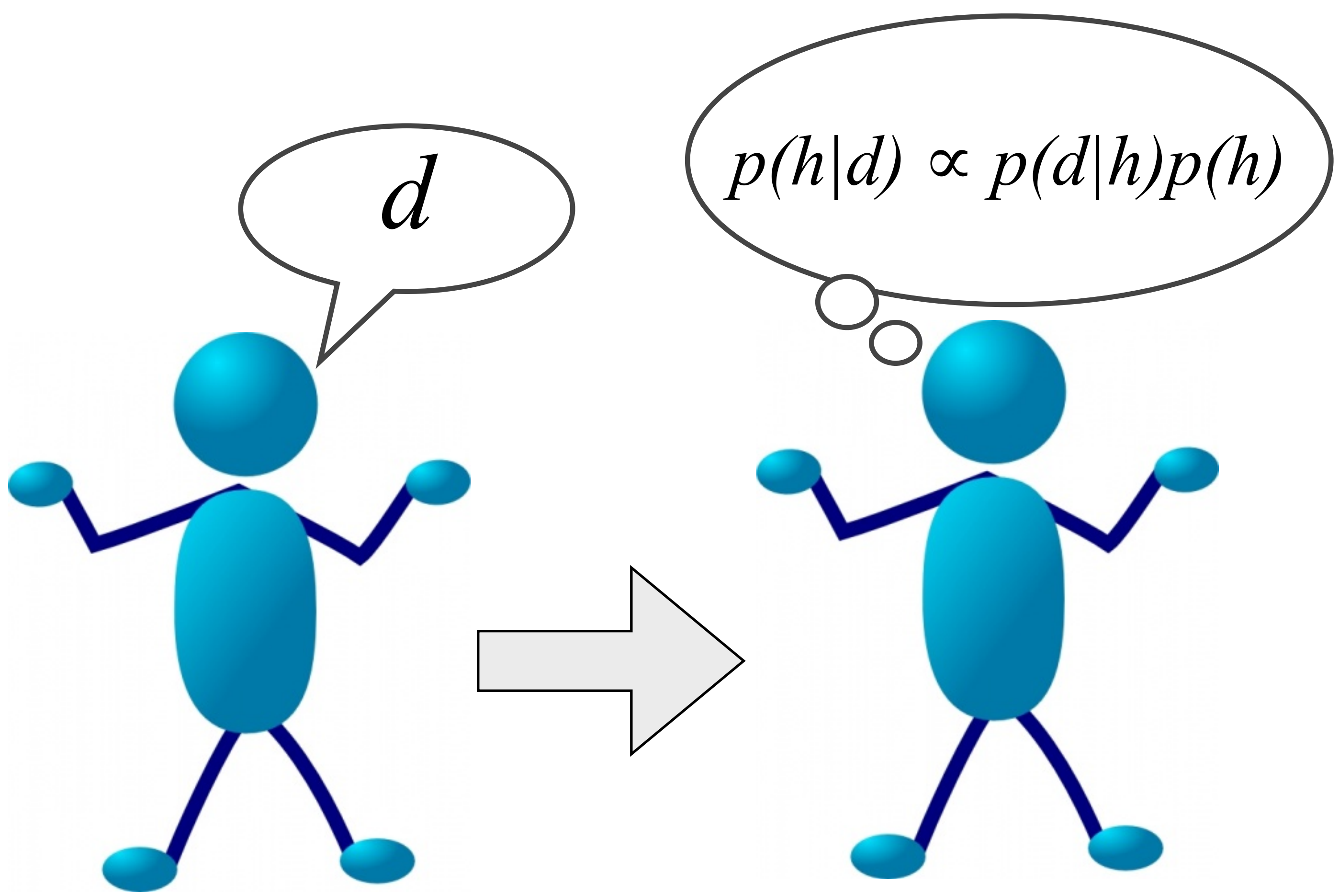
communication dynamics: how data is presented / selected



Iterated learning model: basic idea

- Two main processes occur during this time that can shape how concepts change

cognition: how people learn from the data they see

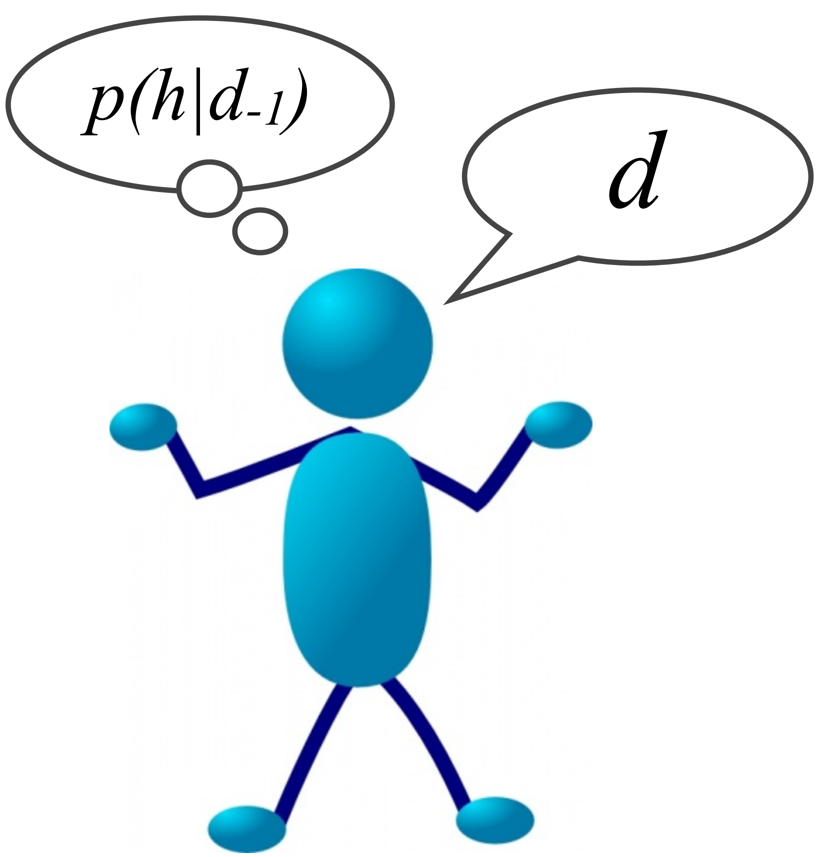


we capture this by assuming that people are Bayesian agents

Iterated learning model: basic idea

- Two main processes occur during this time that can shape how concepts change

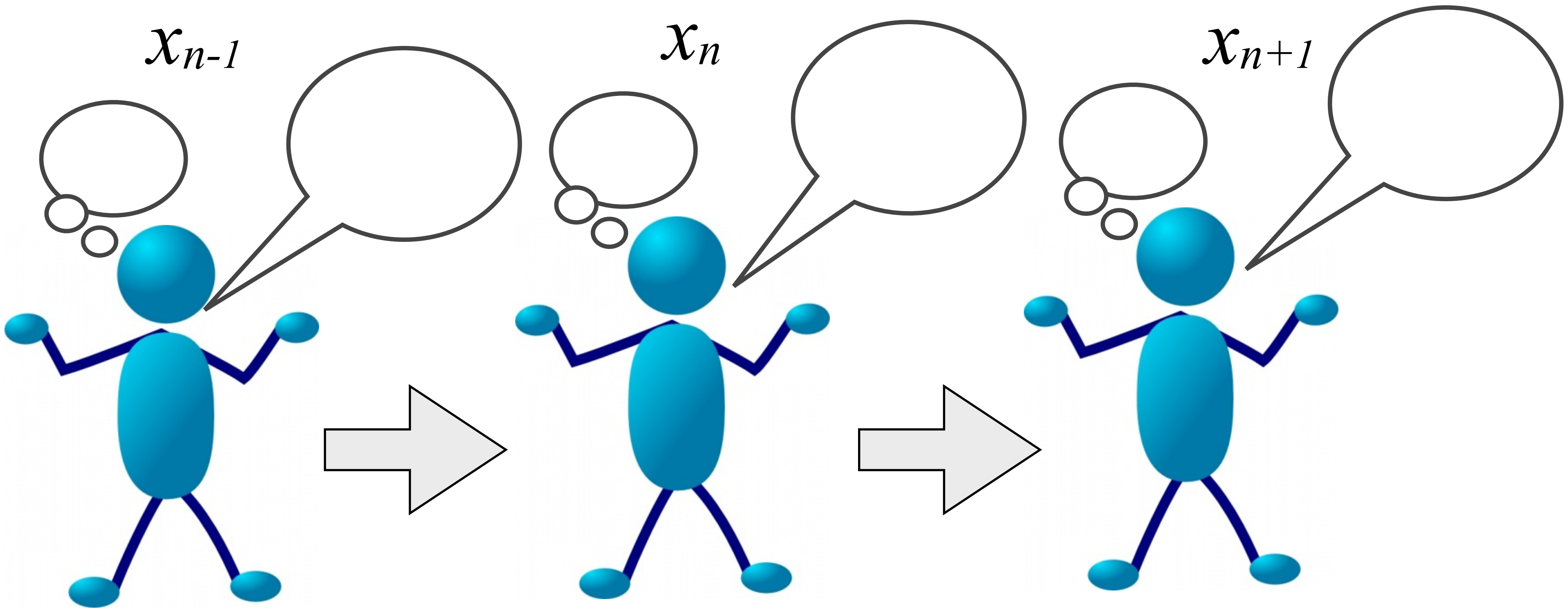
communication dynamics: how data is presented / selected



different results by making different assumptions about how people select the data to transmit and how it might get distorted

to begin with, we'll assume that they just sample a random subset from their inferred distribution over hypotheses

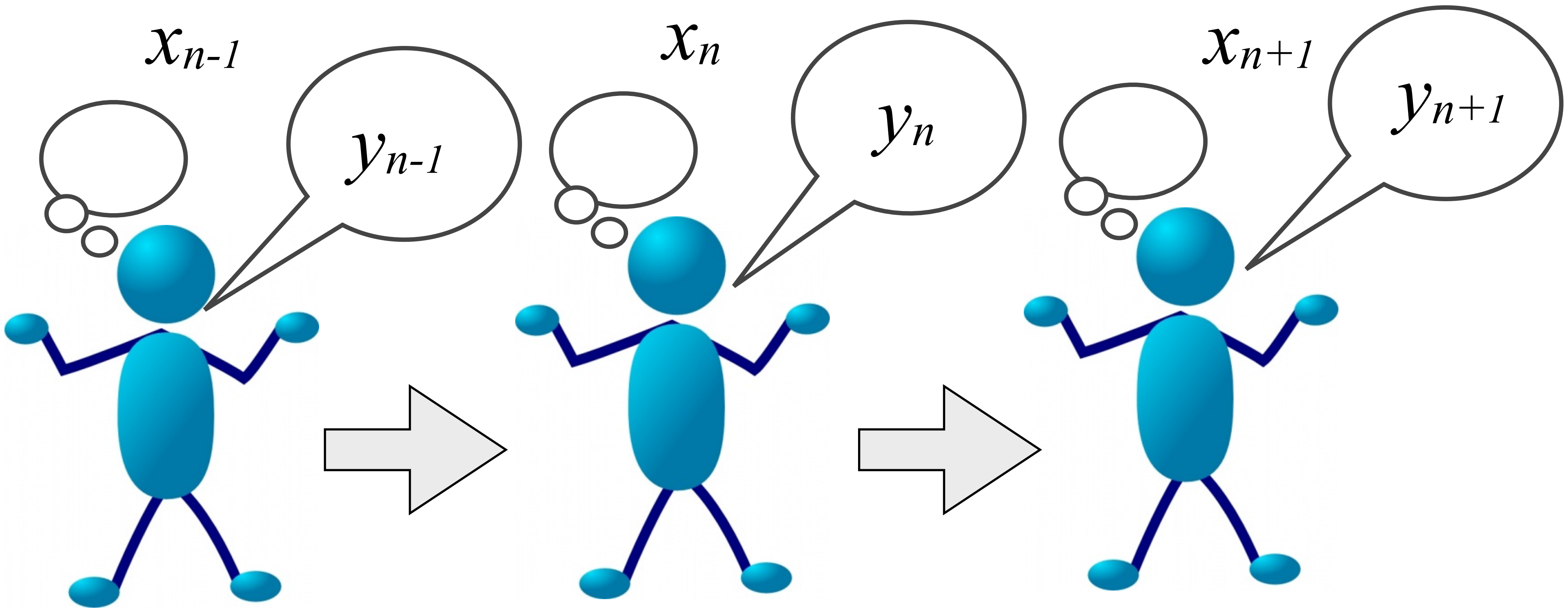
Formalising the model...



x = events in the world

- Cultural or linguistic transmission takes place in world with events x, which are generated from some independent distribution $Q(x)$

Formalising the model...

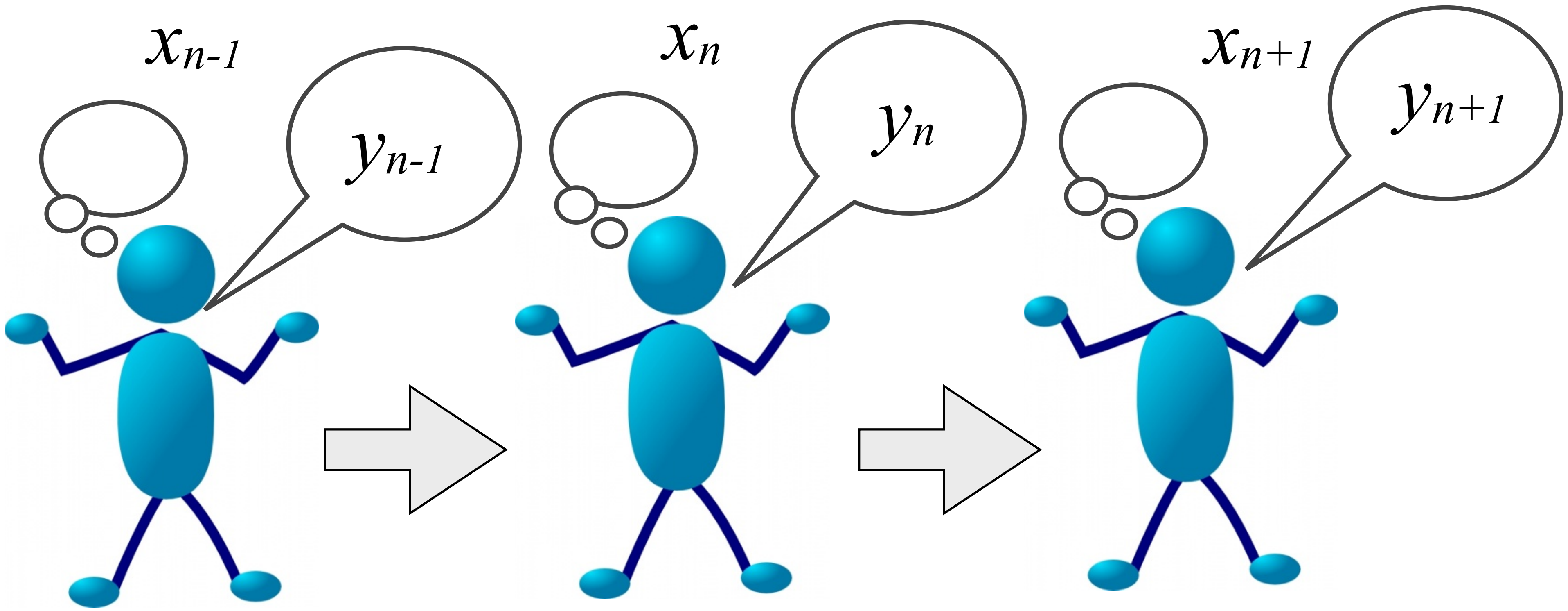


x = events in the world

y = utterance paired with x

- Cultural or linguistic transmission takes place in world with events x, which are generated from some independent distribution $Q(x)$
- For an event x the agent produces some utterance y

Formalising the model...

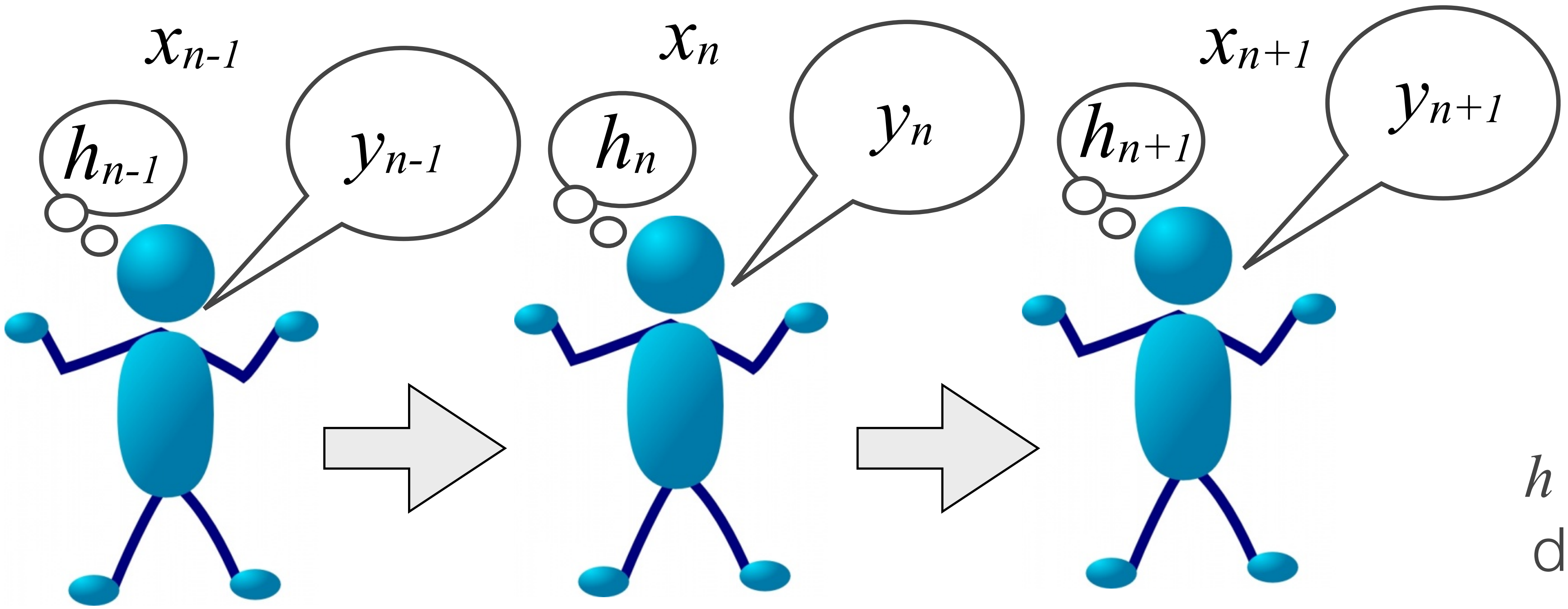


x = events in the world

y = utterance paired with x

- Cultural or linguistic transmission takes place in world with events x, which are generated from some independent distribution $Q(x)$
- For an event x the agent produces some utterance y
- A language/concept is a probability distribution over y for every possible x

Formalising the model...



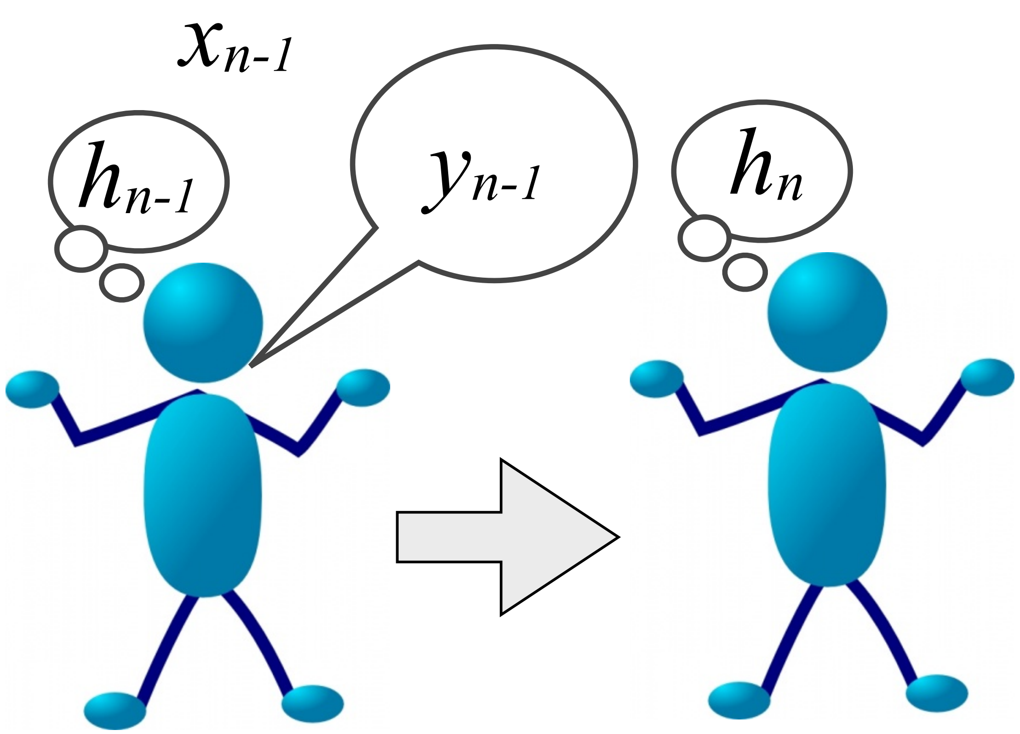
x = events in the world

y = utterance paired with x

h = hypothesis about distributions over x,y

- Cultural or linguistic transmission takes place in world with events x, which are generated from some independent distribution $Q(x)$
- For an event x the agent produces some utterance y
- A language/concept is a probability distribution over y for every possible x
- Assume learners have a set of hypotheses h about the possible concepts

Transmission works like this....



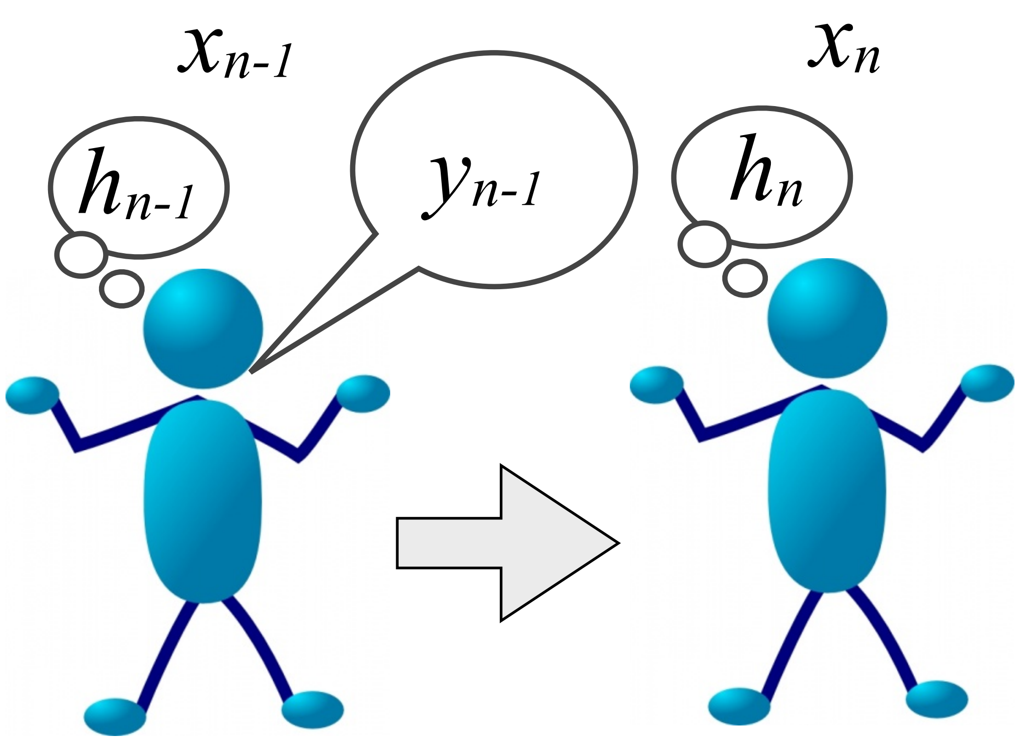
x = events in the world

y = utterance paired with x

h = hypothesis about distributions over x,y

- Previous learners create the input for the next learners

Transmission works like this....



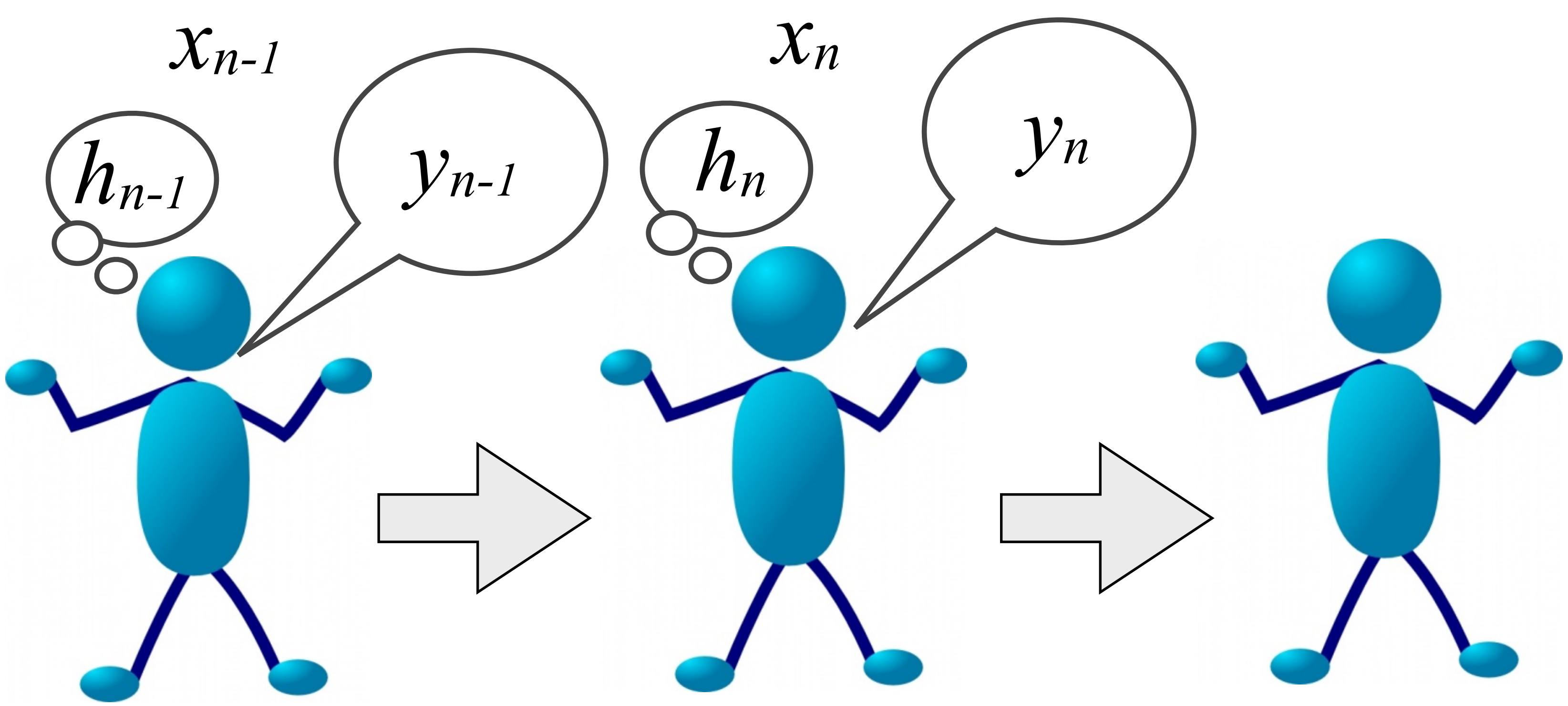
x = events in the world

y = utterance paired with x

h = hypothesis about distributions over x,y

- Previous learners create the input for the next learners
- At time (or for person) n, a series of events x_n occurs

Transmission works like this....



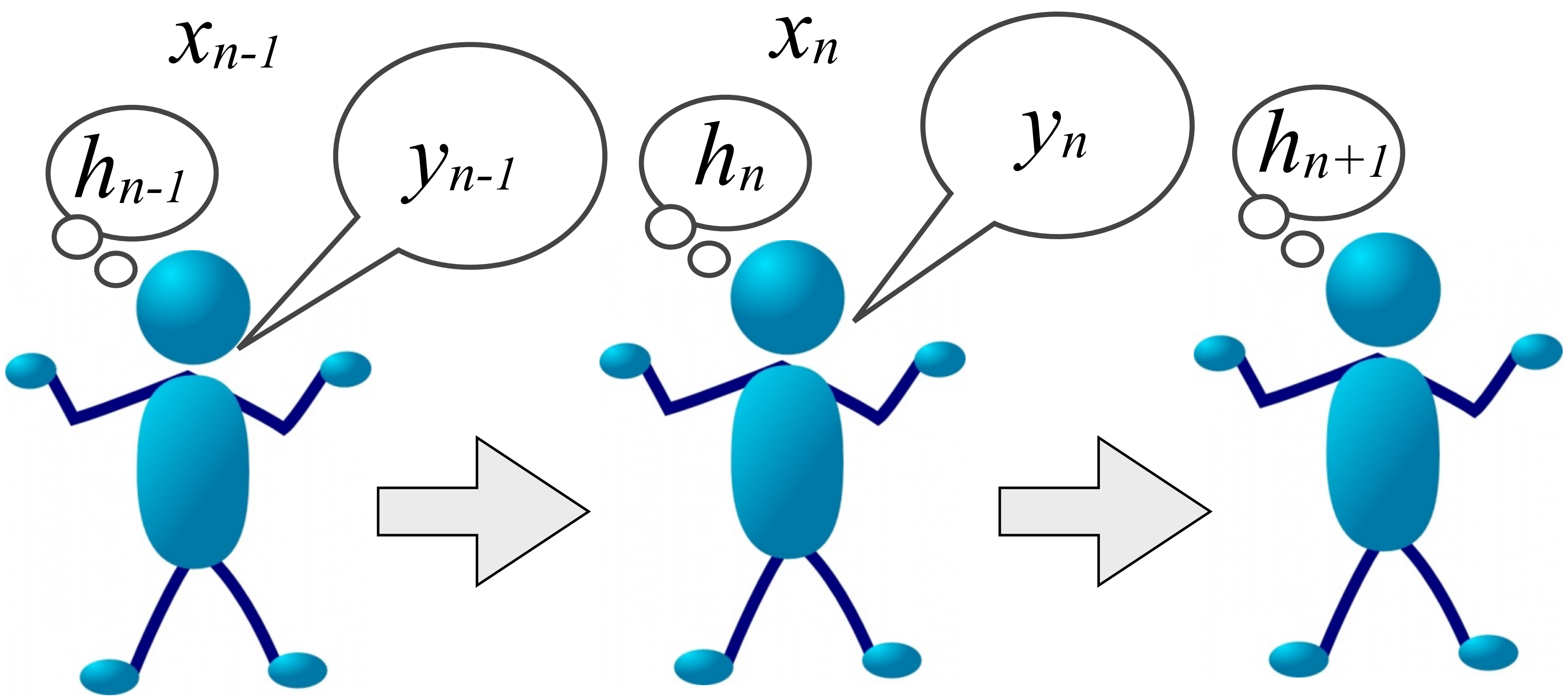
x = events in the world

y = utterance paired with x

h = hypothesis about distributions over x,y

- Previous learners create the input for the next learners
- At time (or for person) n, a series of events x_n occurs
- Based on their hypotheses h_n about the concept, person n speaks utterances y_n to describe those x_n to person $n+1$

Transmission works like this....



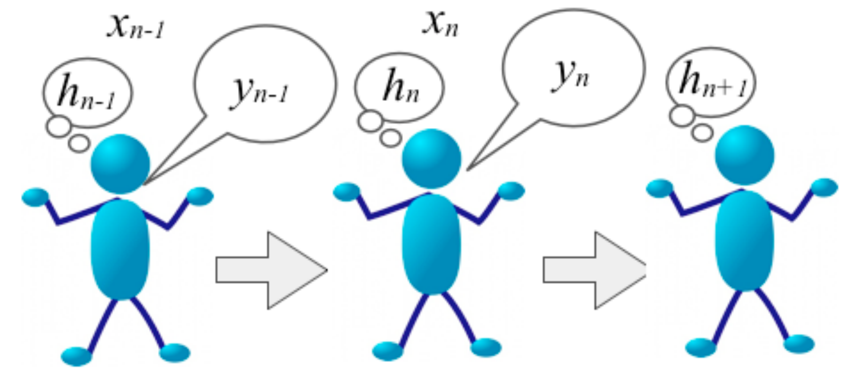
x = events in the world

y = utterance paired with x

h = hypothesis about distributions over x,y

- Previous learners create the input for the next learners
- At time (or for person) n, a series of events x_n occurs
- Based on their hypotheses h_n about the concept, person n speaks utterances y_n to describe those x_n to person $n+1$
- Person $n+1$ infers a hypothesis about the concept based on the data (x,y)

Two steps



- **Learning step**: learner $n+1$ sees x_n (from previous person) and computes a posterior distribution over h_{n+1} according to Bayes' Rule

$$P(h_{n+1}|x_n,y_n) = \frac{P(y_n|x_n,h_{n+1})P(h_{n+1})}{\sum_{h\in\mathcal{H}} P(y_n|x_n,h)P(h)}$$

- **Production step**: Events are generated independently from $Q(x)$. Learner $n+1$ produces utterances y_{n+1} according to

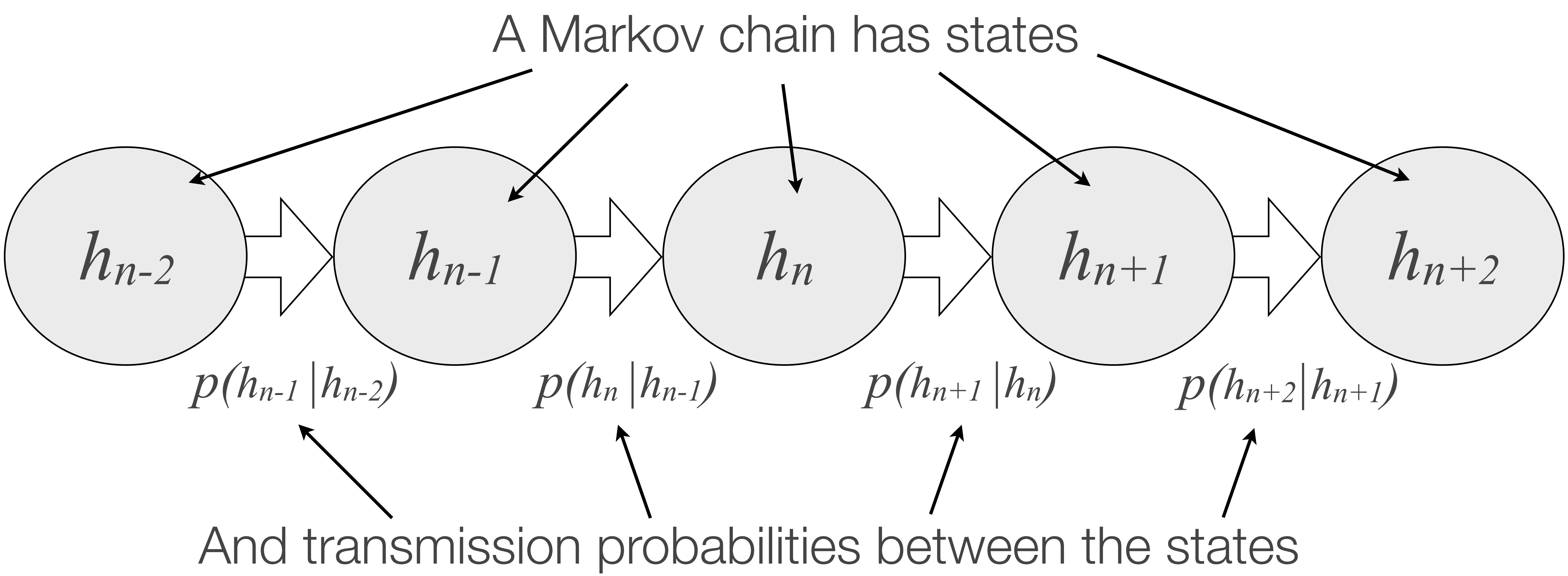
$$P(y_{n+1}|x_{n+1},h_{n+1})$$

- Since all learners use the same learning and production steps, we can calculate:

$$P(h_{n+1}|h_n) = \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} P(h_{n+1}|x,y)P(y|x,h_n)Q(x)$$

This defines a Markov chain

$$P(h_{n+1}|h_n) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P(h_{n+1}|x,y) P(y|x,h_n) Q(x)$$



This defines a Markov chain

Whenever you have states and transmission probabilities between them, you can also write it as a matrix $\mathbf{T}$:

	Red	Blue
Red	$p(R\|R)$	$p(R\|B)$
Blue	$p(B\|R)$	$P(B\|B)$

$\mathbf{T}$

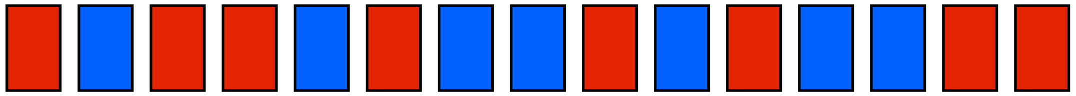
Languages / concepts

	h_1	h_2	h_3
h_1	$p(h_1\|h_1)$	$p(h_1\|h_2)$	$p(h_1\|h_3)$
h_2	$p(h_2\|h_1)$	$p(h_2\|h_2)$	$p(h_2\|h_3)$
h_3	$p(h_3\|h_1)$	$p(h_3\|h_2)$	$p(h_3\|h_3)$

$\mathbf{T}$

This defines a Markov chain

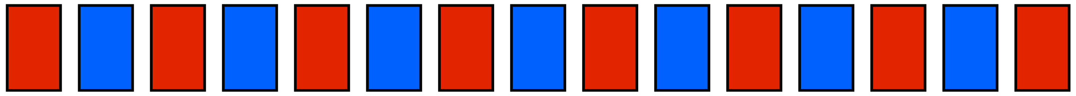
	Red	Blue
Red	0.25	0.75
Blue	0.75	0.25



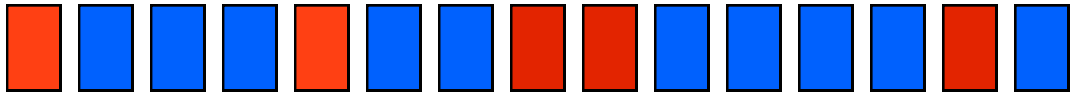
A Markov chain is thus a way of specifying a dynamic process, or a sequence over time

Different chains have different dynamics

	Red	Blue
Red	0	1.0
Blue	1.0	0



	Red	Blue
Red	0.25	0.25
Blue	0.75	0.75



This defines a Markov chain

One way of understanding the dynamics of a chain is to look for "fixed points"

$\mathbf{T}$	Red	Blue
Red	0	1.0
Blue	1.0	0

The stationary distribution π of a Markov chain with transition matrix $\mathbf{T}$ is a distribution such that

$$\pi = \mathbf{T}\pi$$

Or, in other words, the probability distribution over states at point n is the same as the distribution over states at point $n\text{-}1$.

This is stationary because once it has been reached, the probability of being in a particular state will remain constant.

Example: World with two concepts

	h_1	h_2
h_1	$p(h_1\|h_1)$	$p(h_1\|h_2)$
h_2	$p(h_2\|h_1)$	$p(h_2\|h_2)$

simplify the notation:

$$\mathbf{T} = \begin{pmatrix} t_{11} & t_{12} \\ t_{21} & t_{22} \end{pmatrix}$$

these represent mistakes

these represent high-fidelity transmissions

$\boldsymbol{\theta}$ is our probability distribution over languages. θ_1 is the probability that $h=1$, θ_2 is the probability that $h=2$.

$\theta_1 = t_{11}\,\theta_1 + t_{12}\,\theta_2$ (from the definition of the stationary distribution)

and after some math

$$\theta_1 = \frac{t_{12}}{t_{12} + t_{21}} \quad \text{and} \quad \theta_2 = \frac{t_{21}}{t_{12} + t_{21}}$$

What does this mean?

$$\frac{t_{12}}{t_{12} + t_{21}} \quad \text{and} \quad \theta_2 = \frac{t_{21}}{t_{12} + t_{21}} \qquad \mathbf{T} = \begin{pmatrix} t_{11} & t_{12} \\ t_{21} & t_{22} \end{pmatrix}$$

The stationary probability of each of the two concepts is determined by the fidelity with which they are transmitted

	Red	Blue
Red	0	1.0
Blue	1.0	0

	Red	Blue
Red	0.25	0.25
Blue	0.75	0.75

	Red	Blue
Red	0.25	0.75
Blue	0.75	0.25

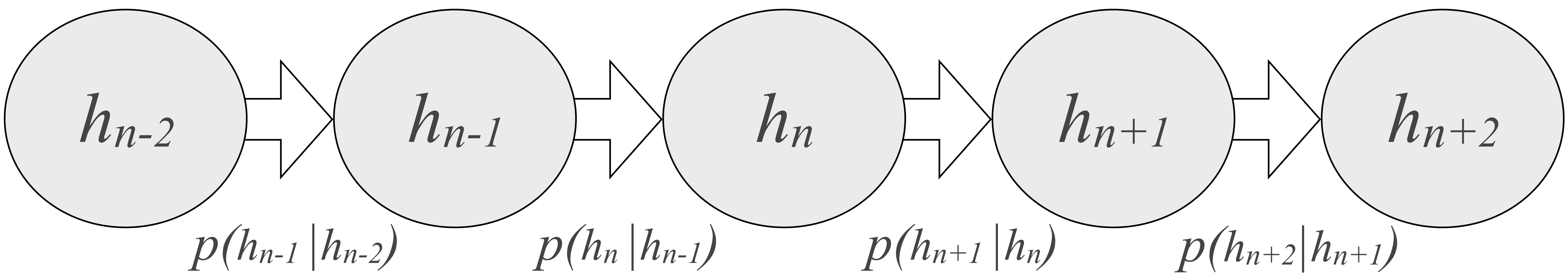
$p(R) = 1/2$
$p(B) = 1/2$

$p(R) = 1/4$
$p(B) = 3/4$

$p(R) = 1/2$
$p(B) = 1/2$

What does this mean about cultural transmission?

Remember we showed that the process of transmission corresponded to a Markov chain over the distribution of concepts or languages

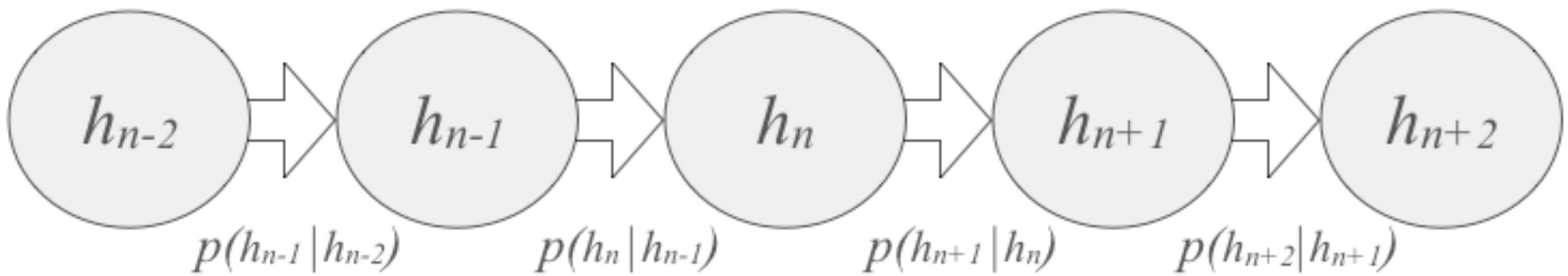


$$P(h_{n+1}|h_n) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P(h_{n+1}|x,y) P(y|x,h_n) Q(x)$$

We can ask what the stationary distribution of this chain is!

This will tell us what distribution of concepts/languages we expect to emerge over time... i.e., which ones will be very frequent and which ones won't be

What does this mean about cultural transmission?

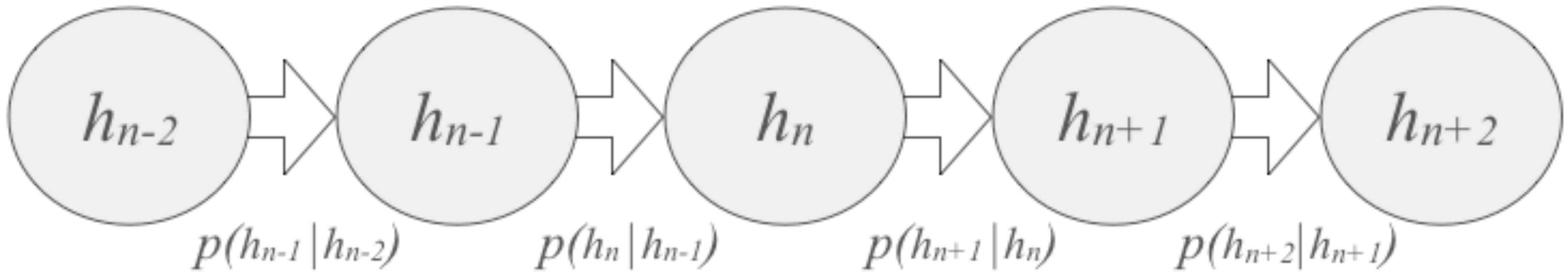


$$P(h_{n+1}|h_n) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P(h_{n+1}|x,y)P(y|x,h_n)Q(x)$$

After a bunch of math, we can prove that the stationary distribution converges to *P(h)*

$$\begin{aligned}
p(h_{n+1}) &= \sum_{h_n \in \mathcal{H}} \left[\sum_{\mathbf{x} \in \mathcal{X}} \sum_{\mathbf{y} \in \mathcal{Y}} p(h_{n+1}|\mathbf{x},\mathbf{y}) p(\mathbf{y}|\mathbf{x},h_n) q(\mathbf{x}) \right] p(h_n) \\
&= \sum_{\mathbf{x} \in \mathcal{X}} \sum_{\mathbf{y} \in \mathcal{Y}} p(h_{n+1}|\mathbf{x},\mathbf{y}) \left[\sum_{h_n \in \mathcal{H}} p(\mathbf{y}|\mathbf{x},h_n) p(h_n) \right] q(\mathbf{x}) \\
&= \sum_{\mathbf{x} \in \mathcal{X}} \sum_{\mathbf{y} \in \mathcal{Y}} \frac{p(\mathbf{y}|\mathbf{x},h_{n+1}) p(h_{n+1})}{p(\mathbf{y}|\mathbf{x})} p(\mathbf{y}|\mathbf{x}) q(\mathbf{x}) \\
&= p(h_{n+1}) \sum_{\mathbf{x} \in \mathcal{X}} \left[\sum_{\mathbf{y} \in \mathcal{Y}} p(\mathbf{y}|\mathbf{x},h_{n+1}) \right] q(\mathbf{x}).
\end{aligned}$$

What does this mean about cultural transmission?

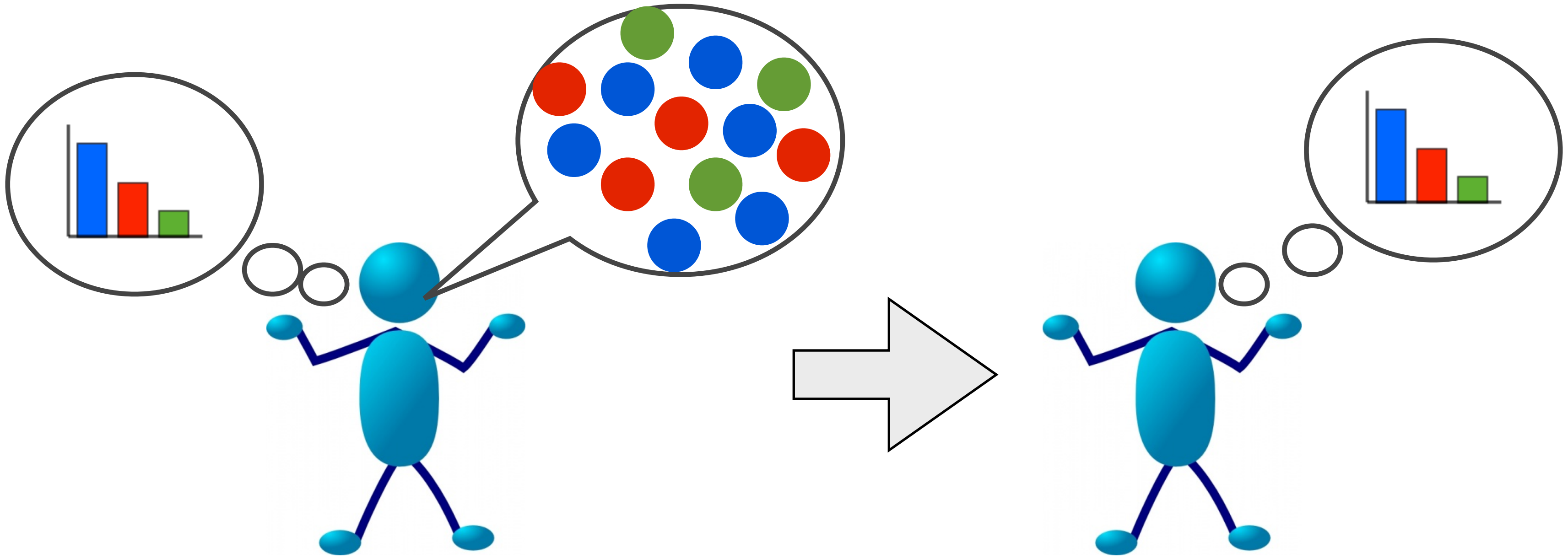


$$P(h_{n+1}|h_n) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P(h_{n+1}|x,y)P(y|x,h_n)Q(x)$$

After a bunch of math, we can prove that the stationary distribution converges to $P(h)$

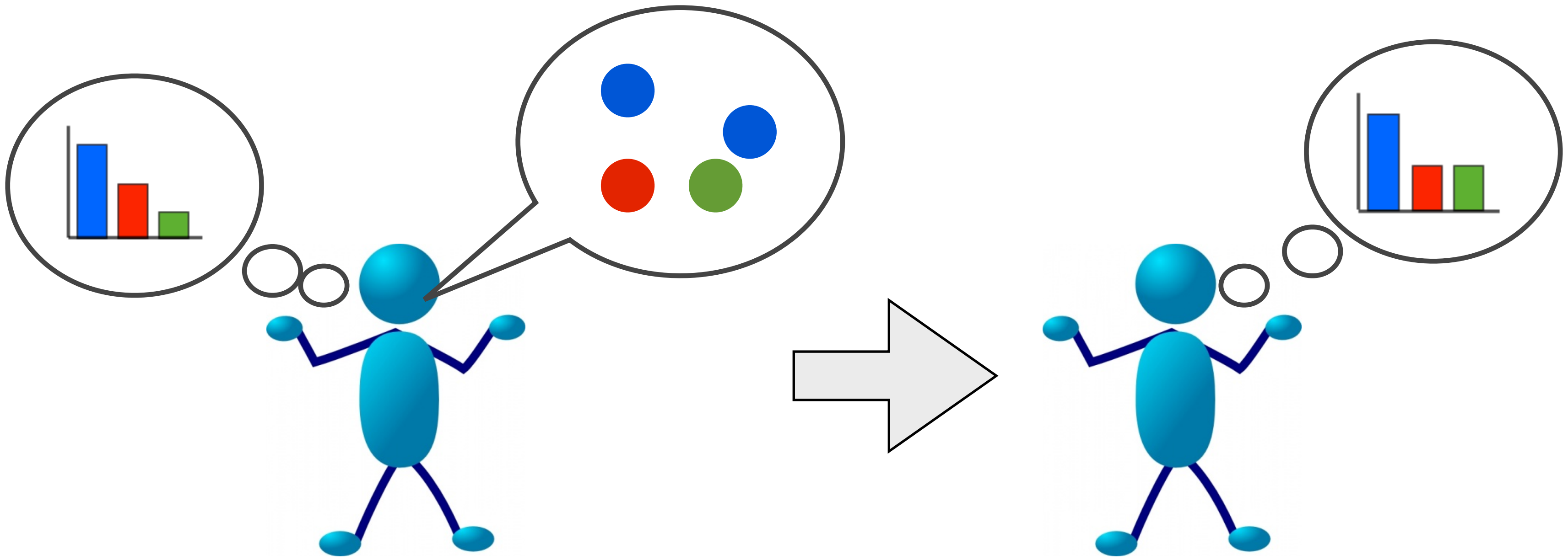
This makes the surprising point that language or cultural transmission / evolution will converge, over time, to people's prior beliefs about the distribution of all possible languages or concepts!

Why convergence to the prior?



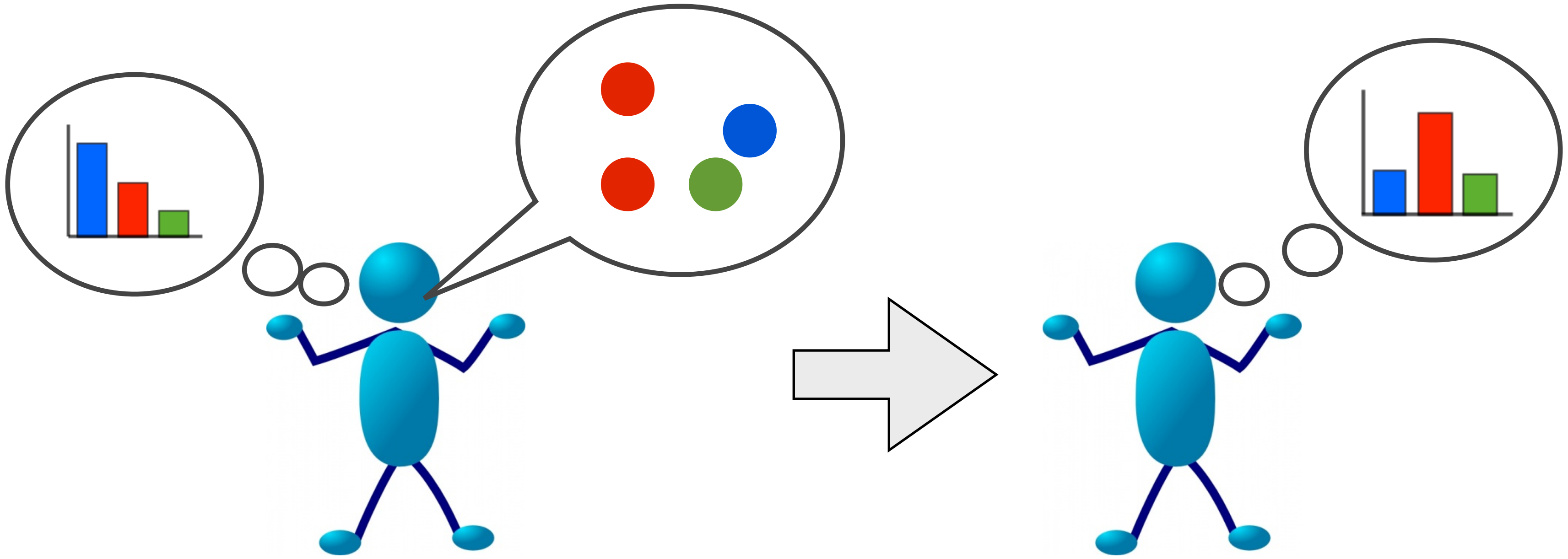
At each point in the chain, it may not be transmitted with perfect fidelity
- bottlenecks mean can only approximate the truth

Why convergence to the prior?



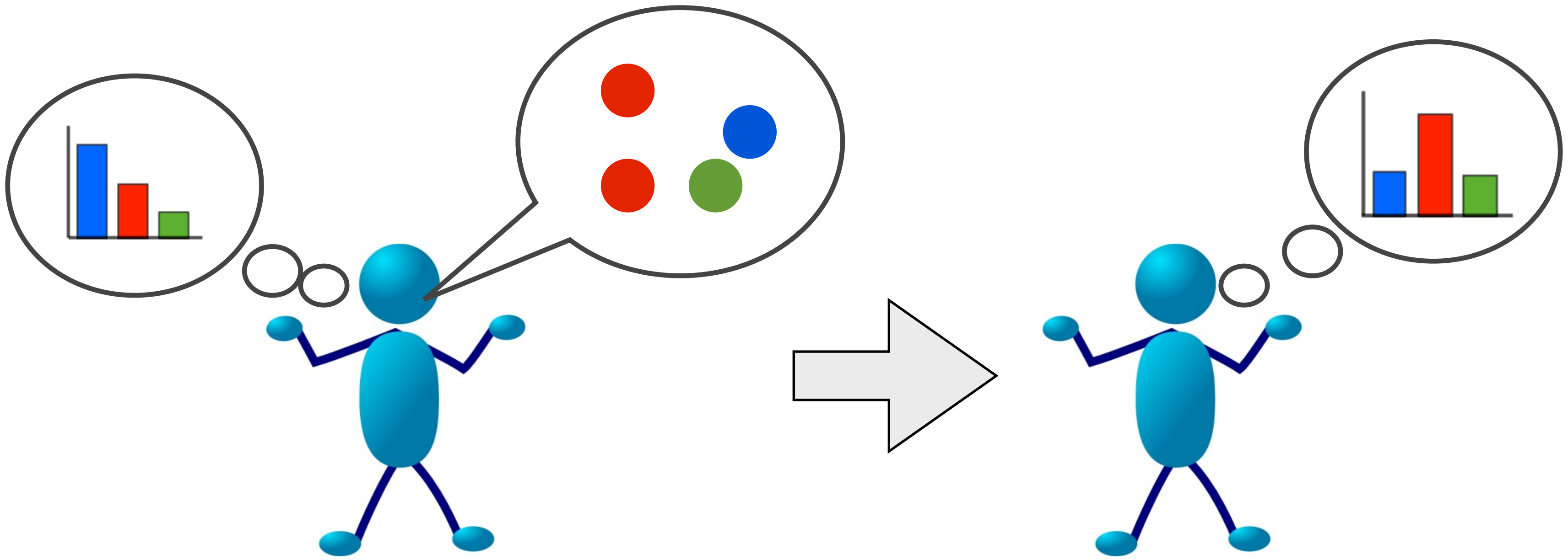
At each point in the chain, it may not be transmitted with perfect fidelity - bottlenecks mean can only approximate the truth

Why convergence to the prior?



At each point in the chain, it may not be transmitted with perfect fidelity
- bottlenecks mean can only approximate the truth
- noise (errors) can make things worse

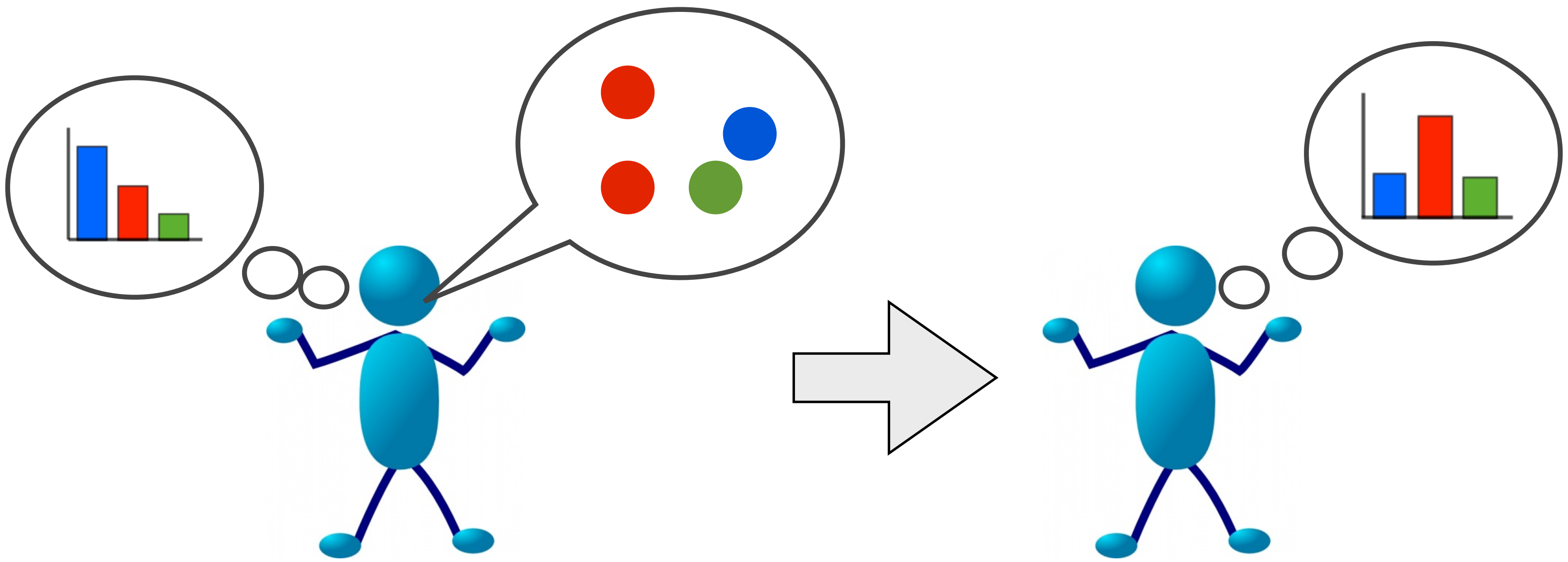
Why convergence to the prior?



At each point in the chain, it may not be transmitted with perfect fidelity
- bottlenecks mean can only approximate the truth
- noise (errors) can make things worse

- When the data are poor, the priors play more of a role

Why convergence to the prior?



At each point in the chain, it may not be transmitted with perfect fidelity
- bottlenecks mean can only approximate the truth
- noise (errors) can make things worse

- When the data are poor, the priors play more of a role
- Over a long time, the initial data are forgotten, and the only stable thing is the prior distribution (which is assumed to be shared)

What does this mean about cultural transmission?

One implication (and test) of this is that if we put people into an iterated-learning-type paradigm, we should see a distribution over their prior beliefs emerge!

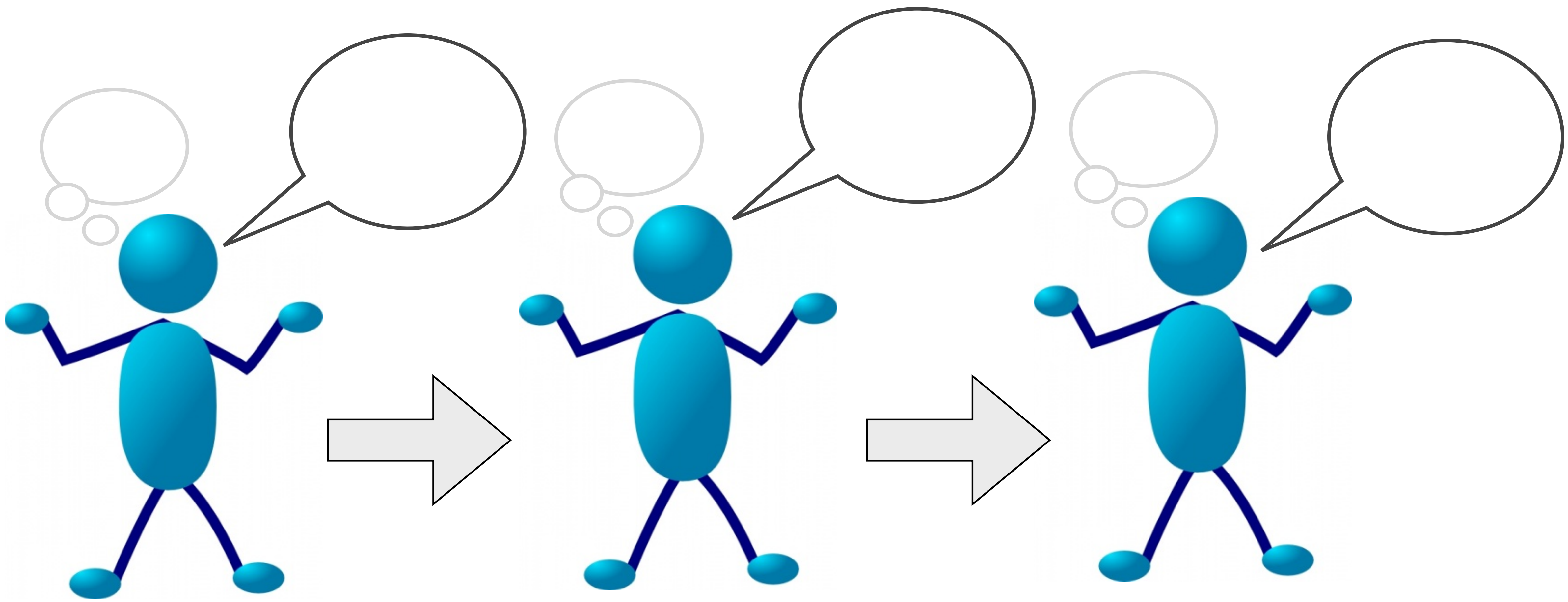
Today's plan

- Evidence for conceptual evolution
  - Inevitable given noisy transmission
  - Historical record
  - Cultural variation
- A model of conceptual change over time
  - Iterated learning model: basic idea
  - Mathematical proof and corresponding intuition
- ➡ Experimental evidence for iterated learning models
  - Function learning
  - Language
- Limitations and extensions to the iterated learning model
  - changing learner
  - changing producer
  - changing how hypotheses map onto the world

Experiments in iterated learning

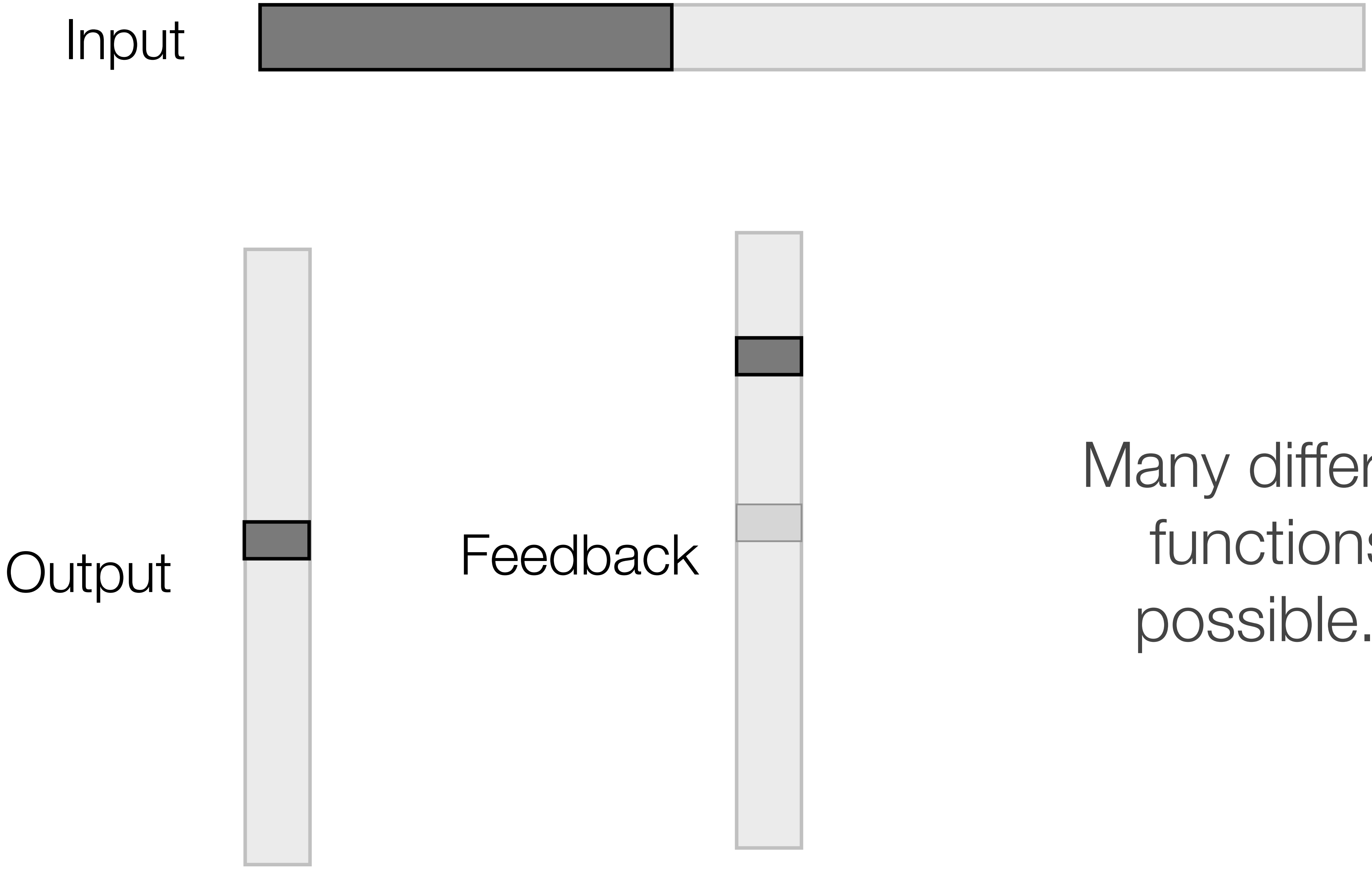
each one produces
some data

that data is given to the next person (not
knowing it came from another participant)



each of these is a
separate participant

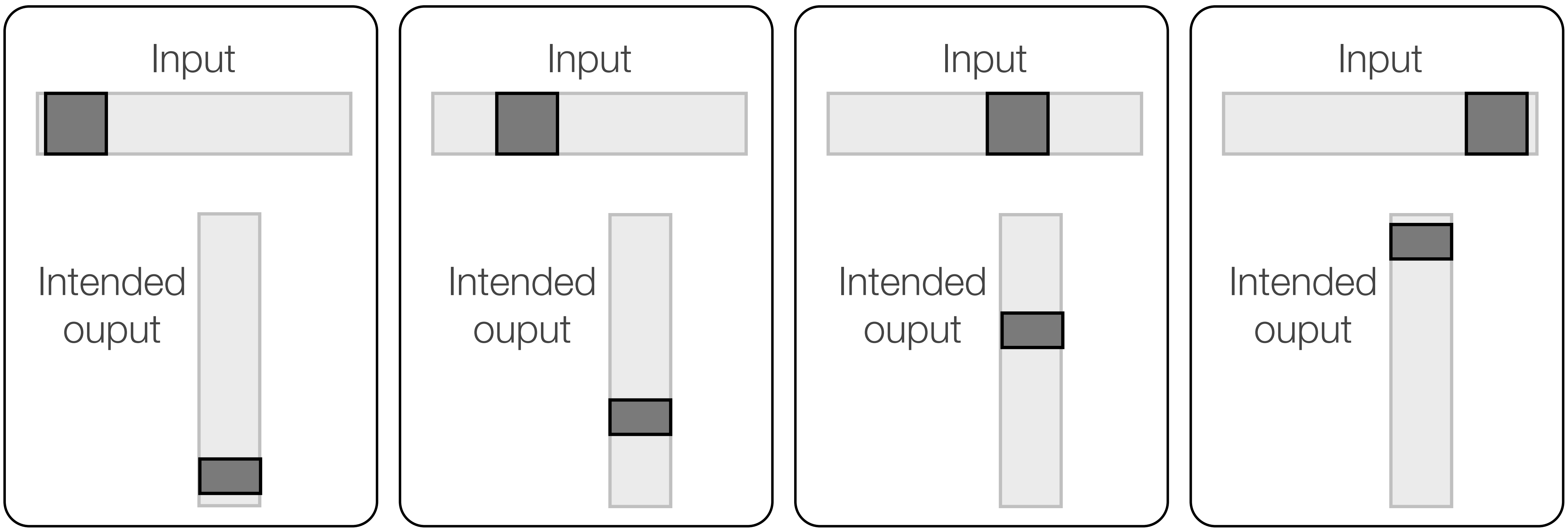
Example #1: Function learning

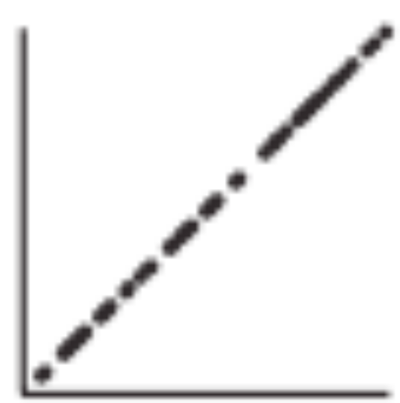


Many different functions possible...

Example #1: Function learning

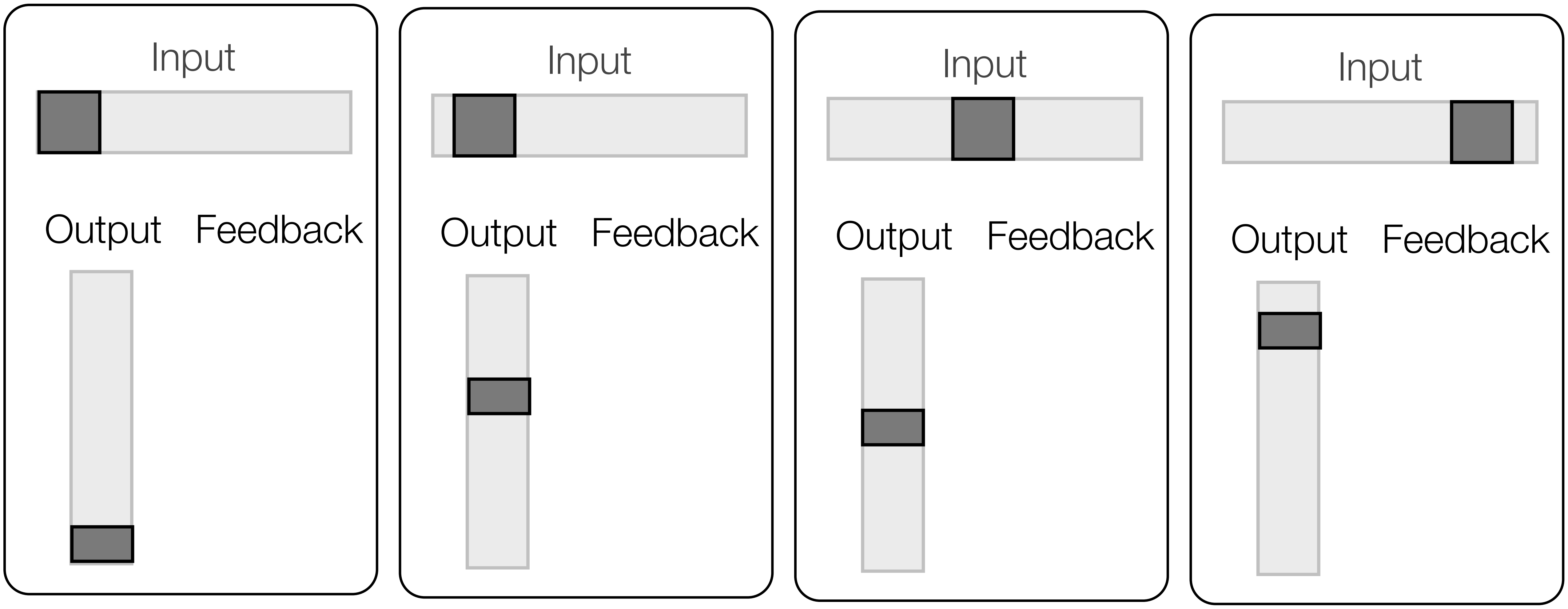
Positive linear function





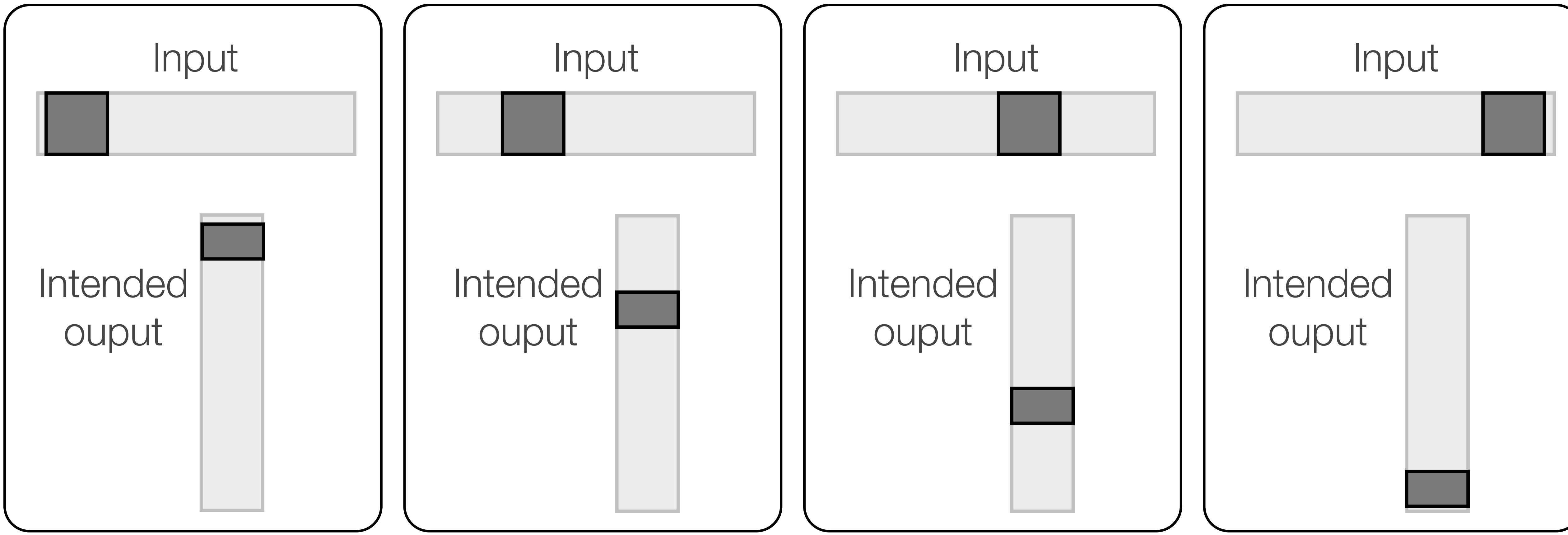
Example #1: Function learning

Positive linear function



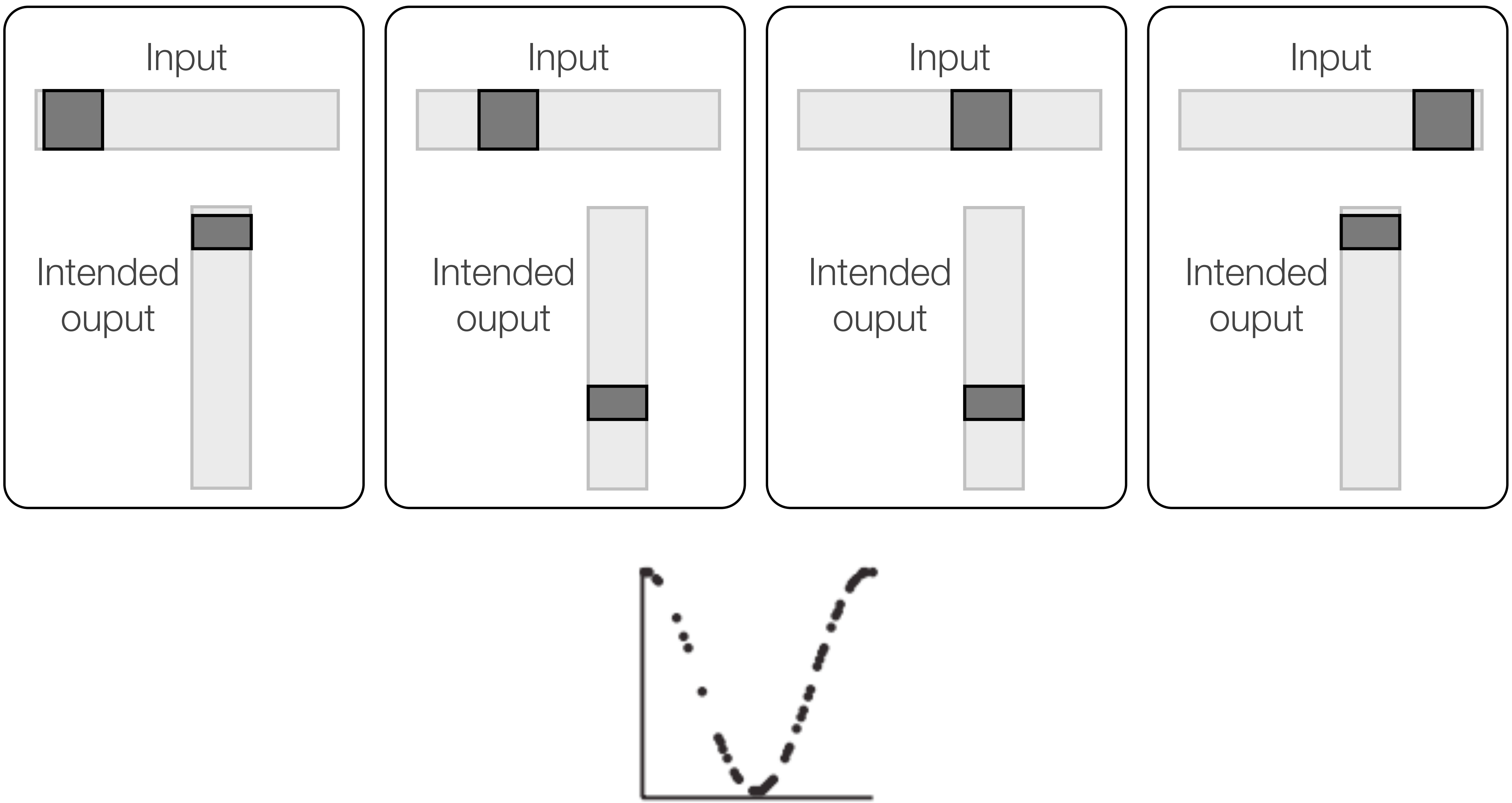
Example #1: Function learning

Negative linear function



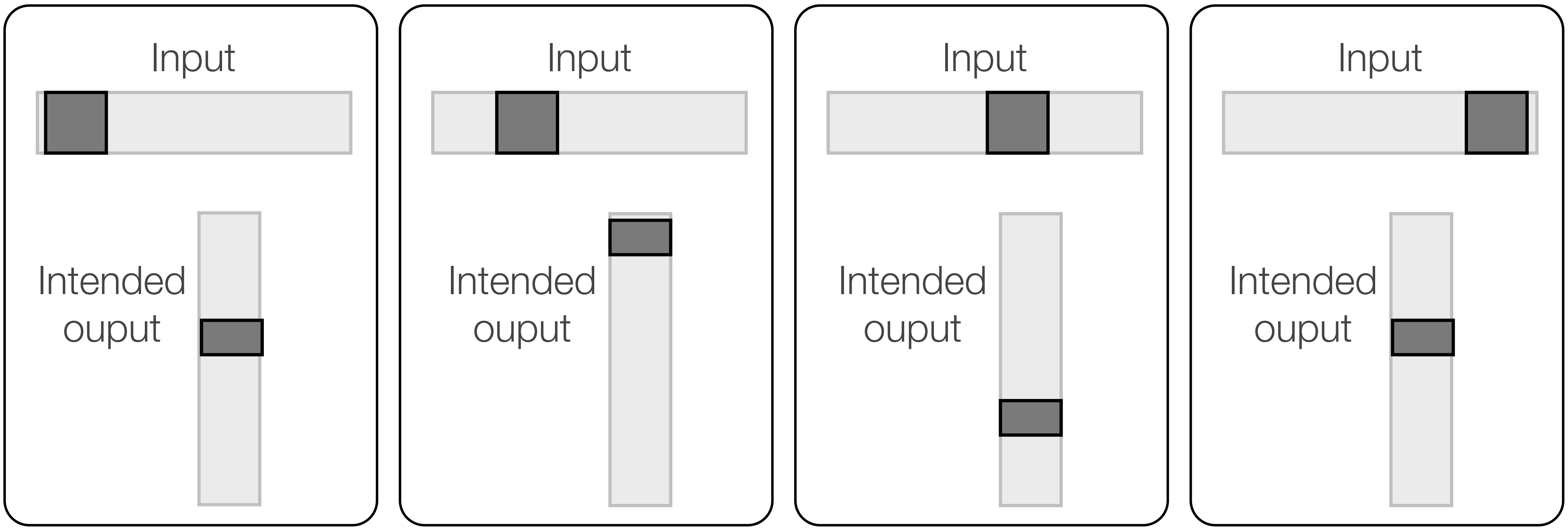
Example #1: Function learning

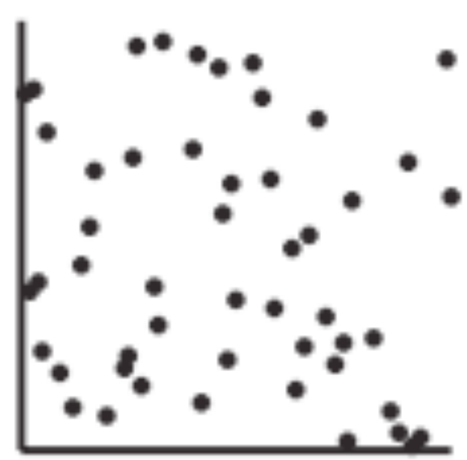
U-shaped function



Example #1: Function learning

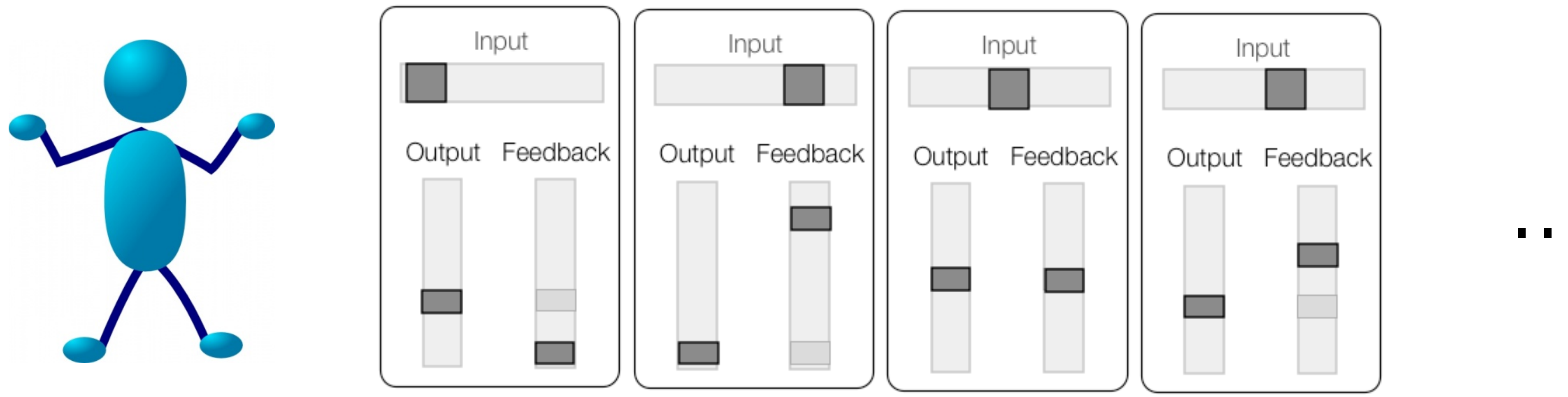
Randomness



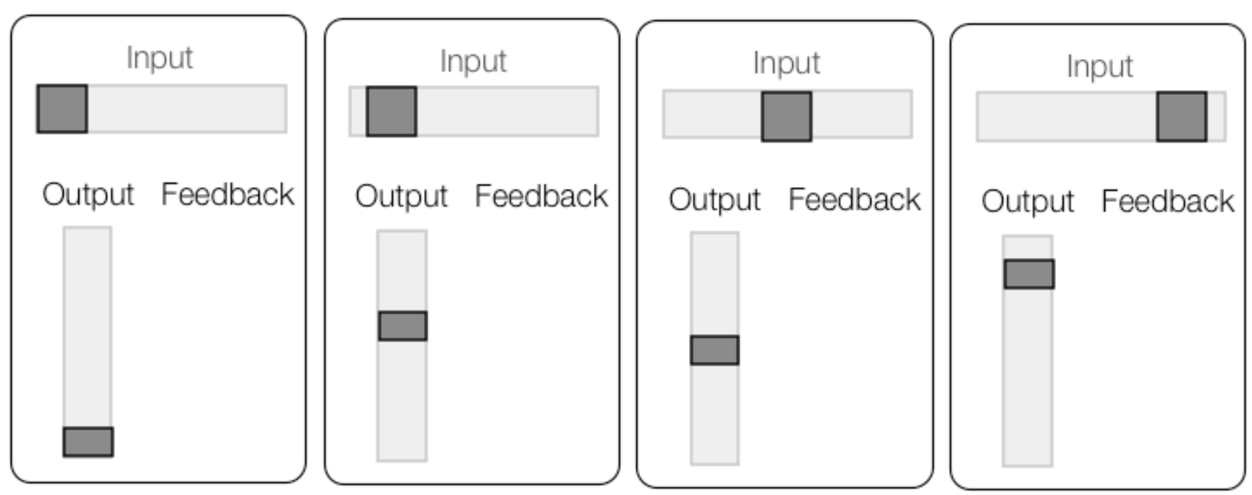


Example #1: Function learning

- First person comes in, gets 50 training items

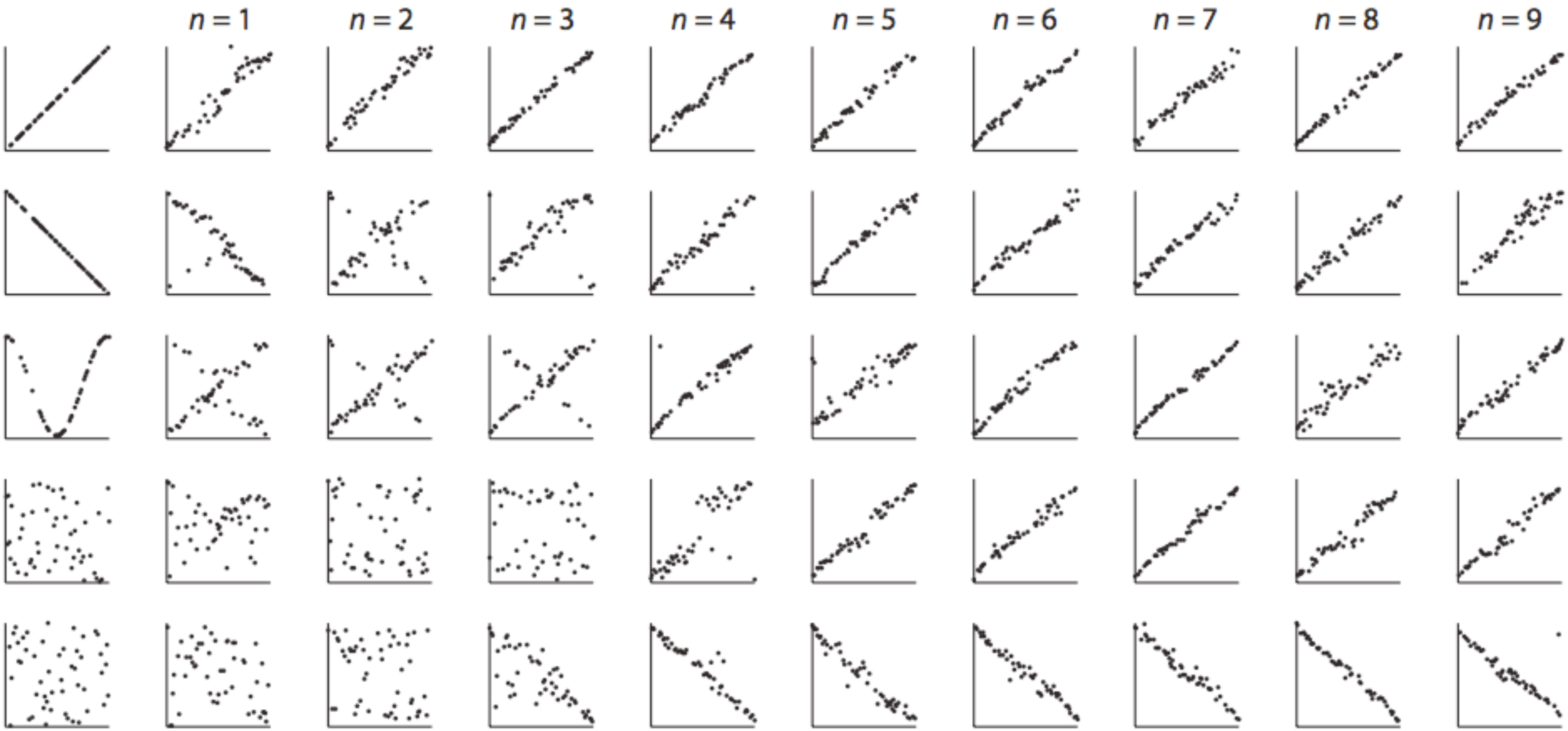


- Then given 25 test items, where they are not given feedback



- These, repeated twice, serve as the next person's training items

Example #1: Function learning



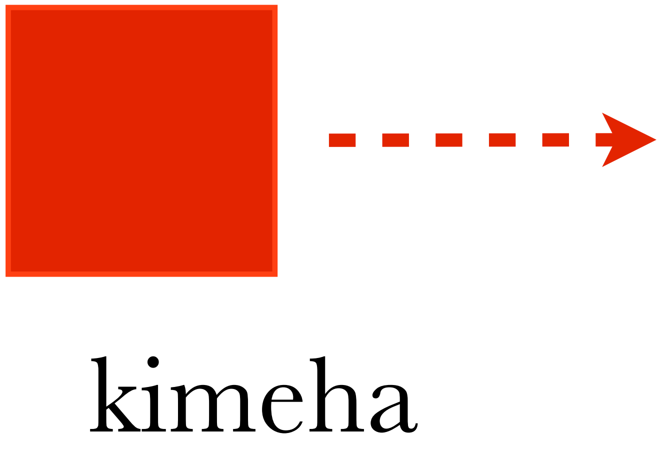
- In nearly all cases the chain converged on a positive linear function!
- Occasionally negative linear, but relatively rare

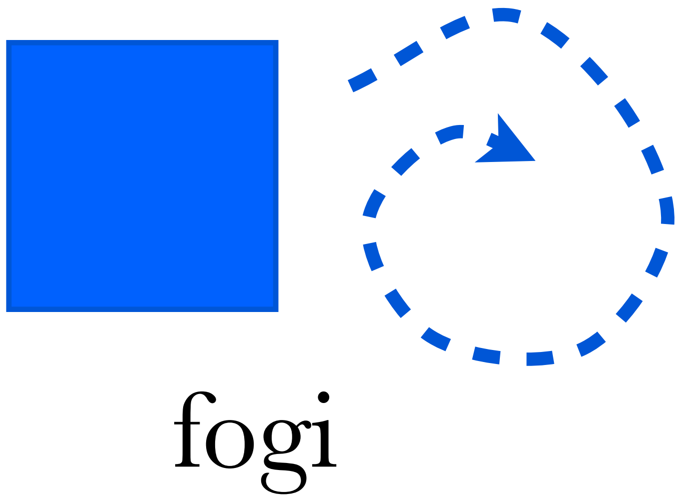
Example #2: Language

- 27 possible “events” in the world
  - Three shapes: square, circle, triangle
  - Three motions: horizontal, bouncing, spiraling
  - Three colours: red, black, blue

Example #2: Language

- 27 possible "events" in the world
  - Three shapes: square, circle, triangle
  - Three motions: horizontal, bouncing, spiraling
  - Three colours: red, black, blue
- Events are paired with a label (for first person, it's random)

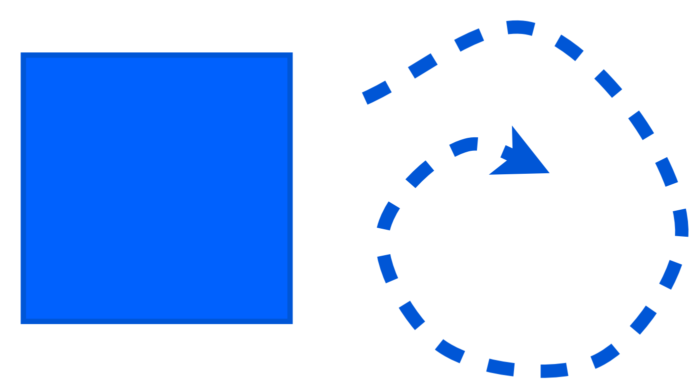




Example #2: Language

- 27 possible “events” in the world
  - Three shapes: square, circle, triangle
  - Three motions: horizontal, bouncing, spiraling
  - Three colours: red, black, blue
- Events are paired with a label (for first person, it’s random)
- After training on these, the person is shown events and has to generate the label themselves

kihema



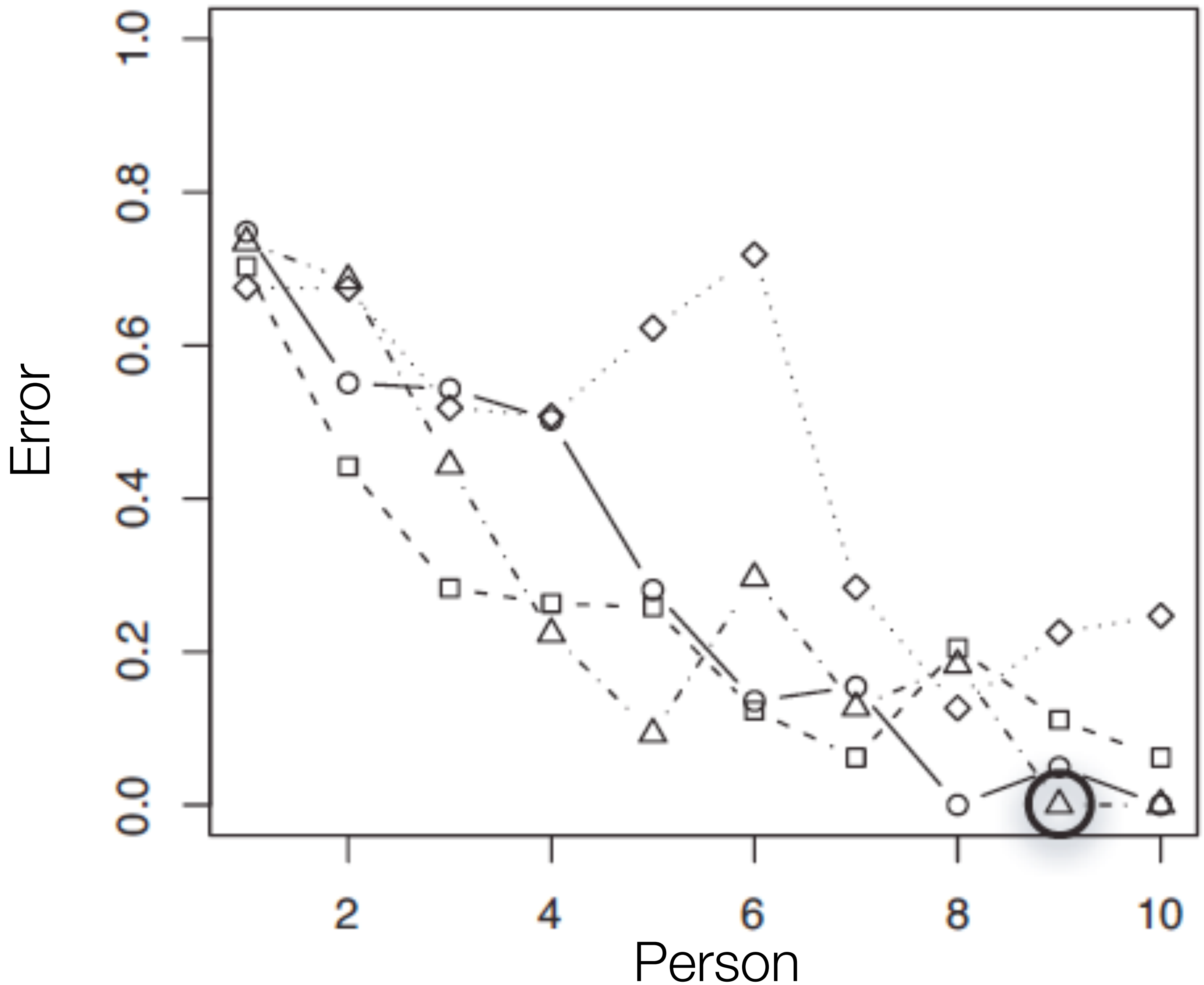
kigi

Example #2: Language

- 27 possible “events” in the world
  - Three shapes: square, circle, triangle
  - Three motions: horizontal, bouncing, spiraling
  - Three colours: red, black, blue
- Events are paired with a label (for first person, it’s random)
- After training on these, the person is shown events and has to generate the label themselves
- The next person is given the previous person’s labels to use as their training data

Example #2: Language

- Transmission error goes down significantly over time



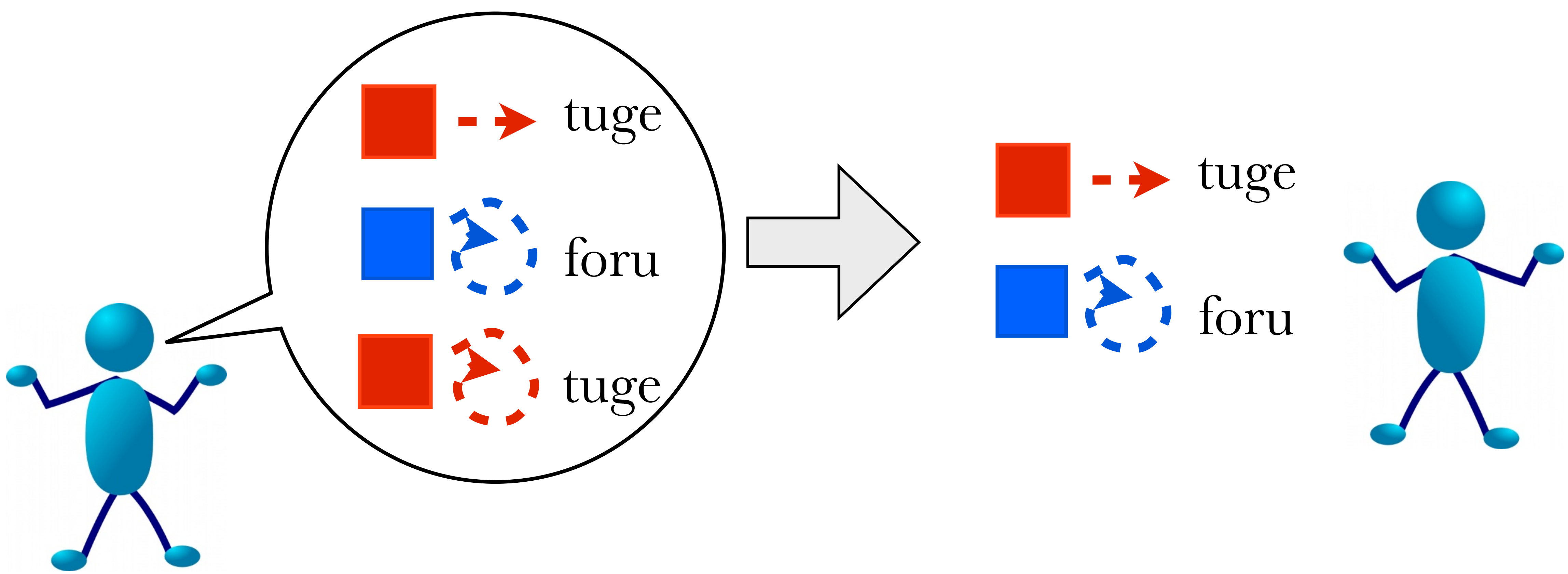
Example #2: Language

- This is achieved by generating languages that are underspecified

Example #2: Language

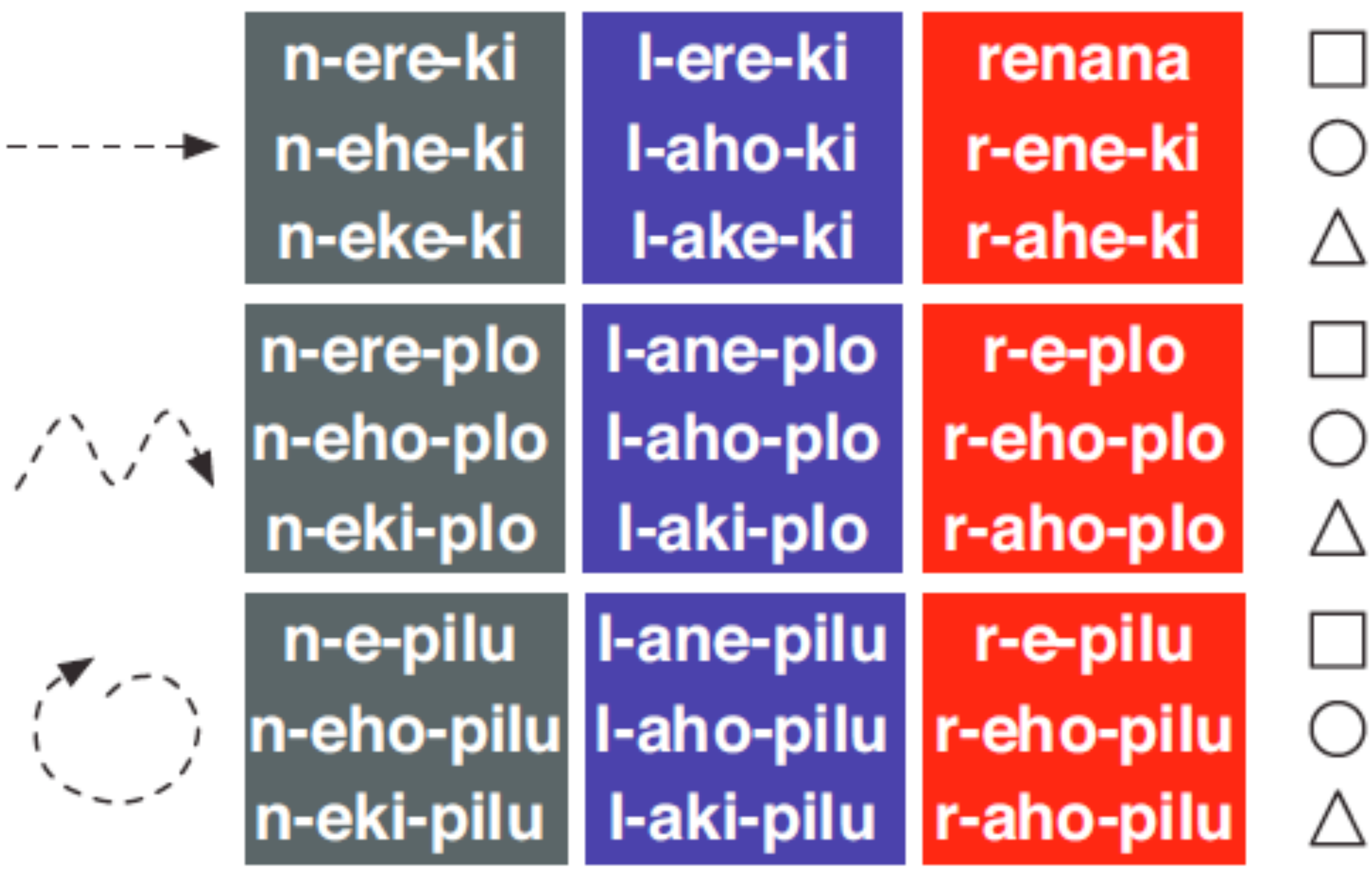
- We can impose a pressure against ambiguity by “filtering” the data between participants

If anyone assigns a string to more than one meaning, all meanings except one (chosen at random) are removed from the next person’s training set



Example #2: Language

- When there is no pressure for ambiguity, compositional language emerges (each word part corresponds to one aspect of meaning)



Summary

- We've started looking a little about how to study things that change over time. As a first stab, we're looking at conceptual change/evolution over time

Ancient Greece/Rome:
- marriages arranged
- affairs (for men) okay, including with young boys
- not usually for love

medieval times:
- still usually economic
- often involved dowries
- women were property
- church involved more

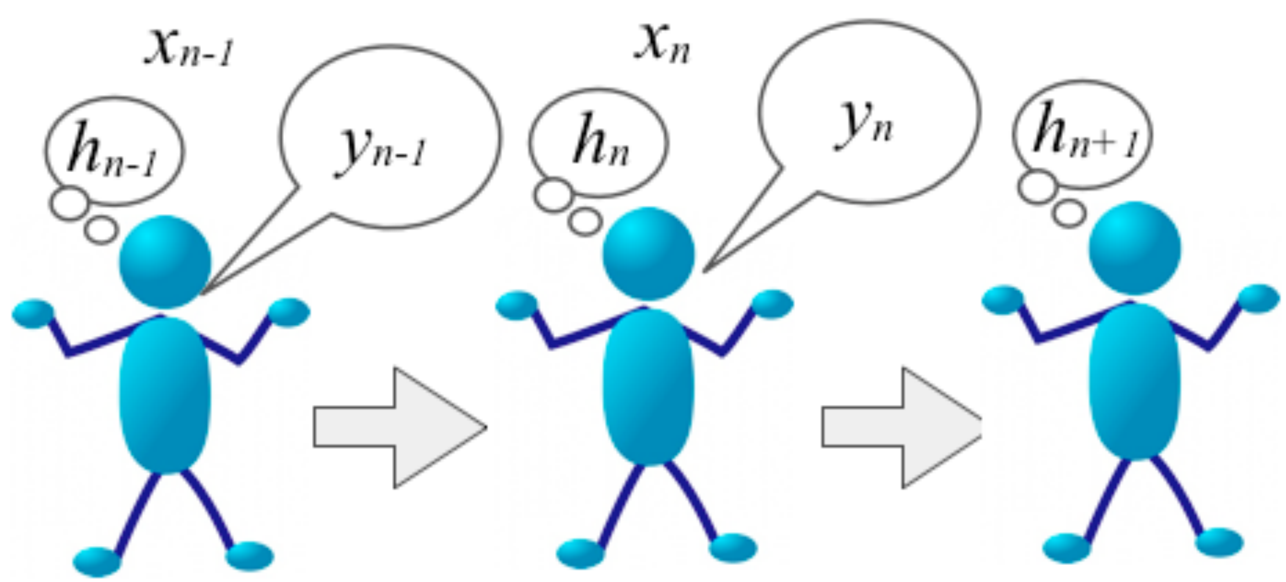
19th century times:
- occasionally for love
- often not cross-racial
- women sometimes can keep property

1950s etc
- often for love
- cross racial sometimes ok
- women "in the home"

today:
- usually for love
- cross-racial okay
- same-sex sometimes ok
- women wield much more economic power

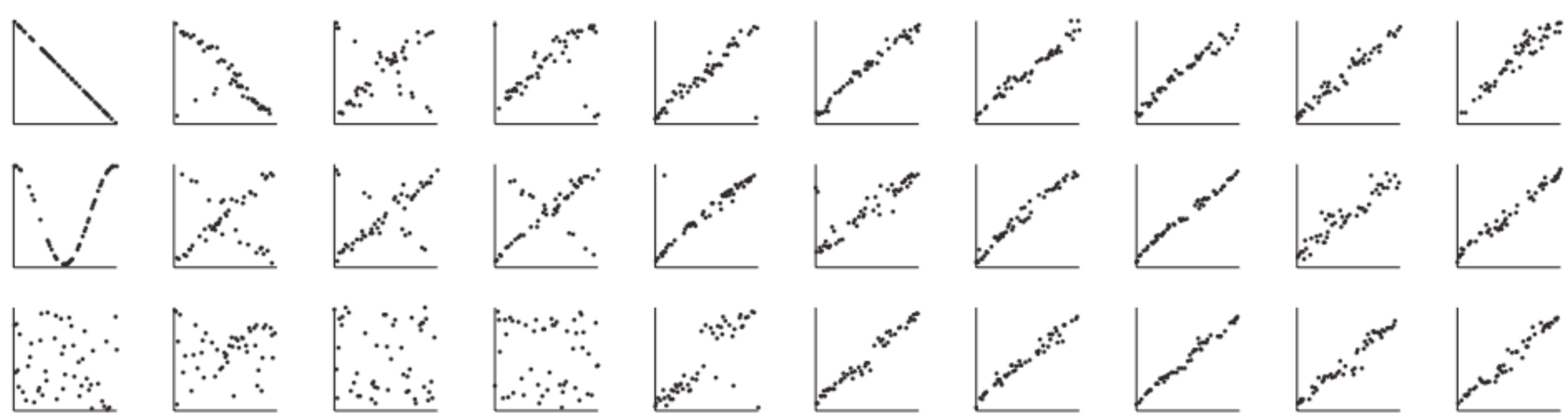
Summary

- We've started looking a little about how to study things that change over time. As a first stab, we're looking at conceptual change /evolution over time
- This is modeled as chains of learners who pass each other information, and are individually Bayesian in how they learn from the previous one



Summary

- We've started looking a little about how to study things that change over time. As a first stab, we're looking at conceptual change /evolution over time
- This is modeled as chains of learners who pass each other information, and are individually Bayesian in how they learn from the previous one
- The main prediction, that the stationary distribution of the chain reflects (only) prior probability, was borne out experimentally



Additional references (not required)

Mathematical analyses

- Griffiths, T., & Kalish, M. (2005). A Bayesian view of language evolution by iterated learning. *Proceedings of the 27th Cognitive Science Conference*.
- Griffiths, T., & Kalish, M. (2007). Language evolution by iterated learning with Bayesian agents. *Cognitive Science*, 31, 441-480.
- Smith, K. (2009). Iterated learning in populations of Bayesian agents. *Proceedings of the 31st Cognitive Science Conference.*
- Perfors, A., & Navarro, D. (2014). Language evolution can be shaped by the structure of the world. *Cognitive Science.*

Experimental results

- Kalish, M., Griffiths, T., & Lewandowsky, S. (2007). Iterated learning: Intergenerational knowledge transmission reveals inductive biases. *Psychonomic Bulletin and Review*, 14(2), 288-294.
- Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences,* 105(31), 10681-10686.
- Beppu, A., & Griffiths, T. (2009). Iterated learning in populations of Bayesian agents. *Proceedings of the 31st Cognitive Science Conference.*