Simulating Language
Lecture 9: Uncertainty during word learning

Simon Kirby

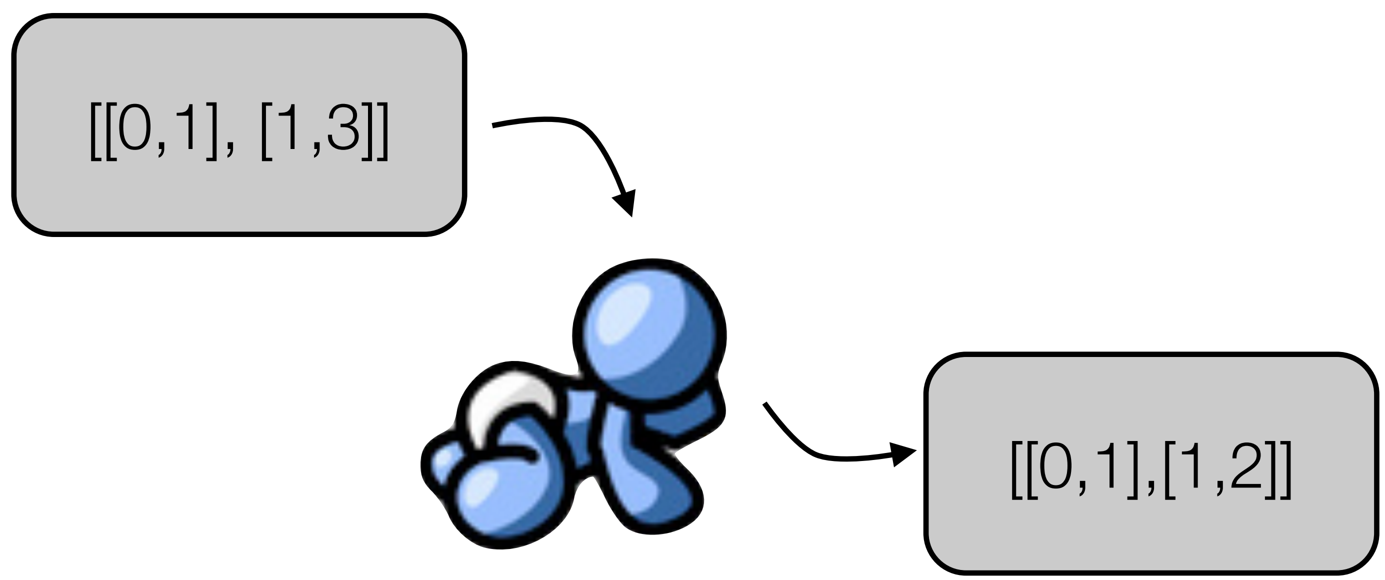
simon@ling.ed.ac.uk

Learning and Evolution

- So far, we've looked at three different processes:
  - Social learning mechanisms
  - Cultural evolution of learnt behaviour
  - Biological evolution (of connection weights, or of learning mechanisms in Smith, 2004)
- Today we're going to go back to learning, and particularly how children learn the meanings of words.

Meaning

- In our models so far, learning has required the explicit presentation of meaning-signal pairs to the learner.



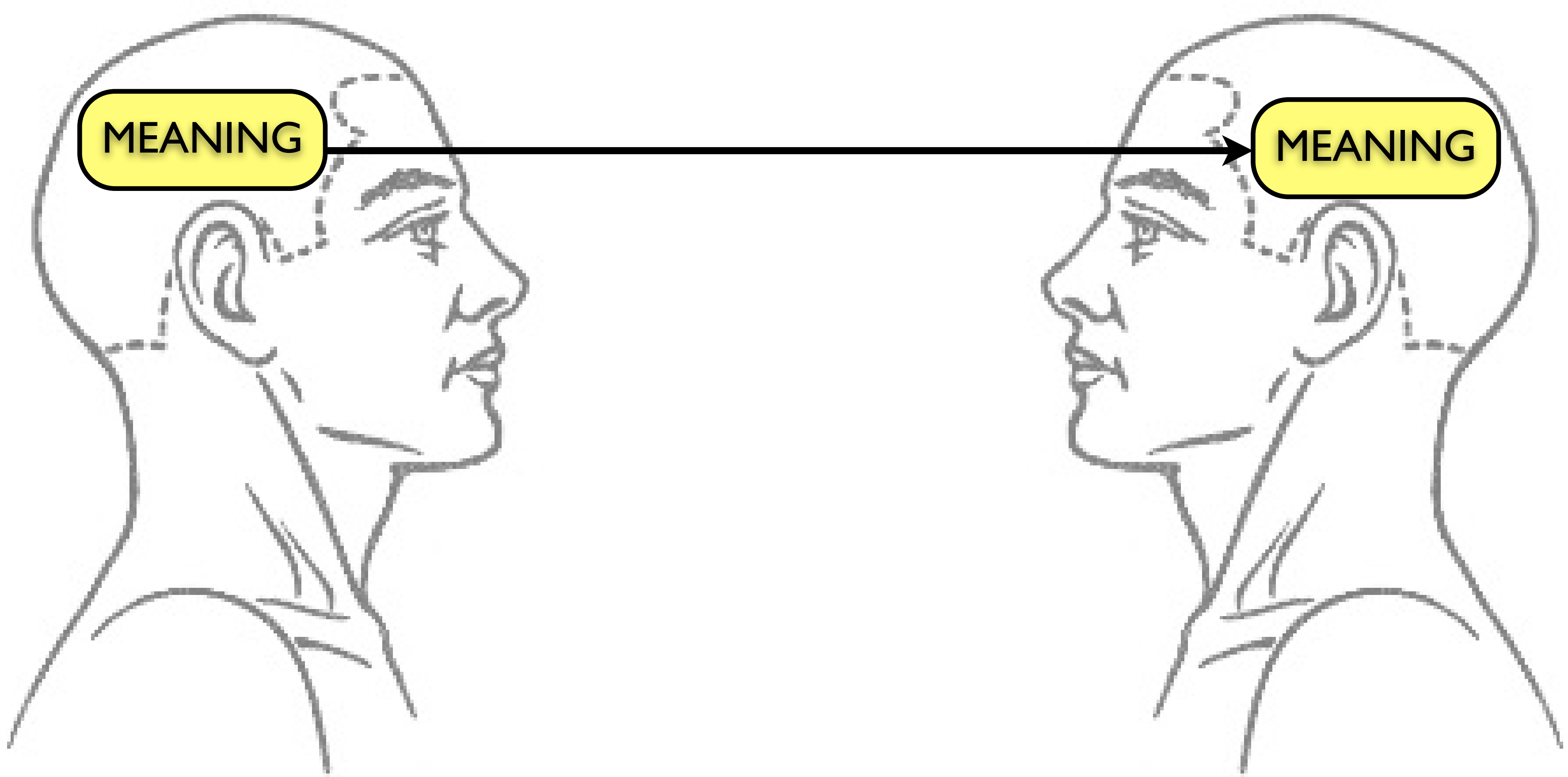
- The learner then outputs another set of meaning-signal pairs for the next generation.

- But are meanings really directly presented to learners?

Communication

- In communication, the speaker informs the hearer about some state of affairs, and this information triggers some response in the hearer (such as a change in their cognitive state).

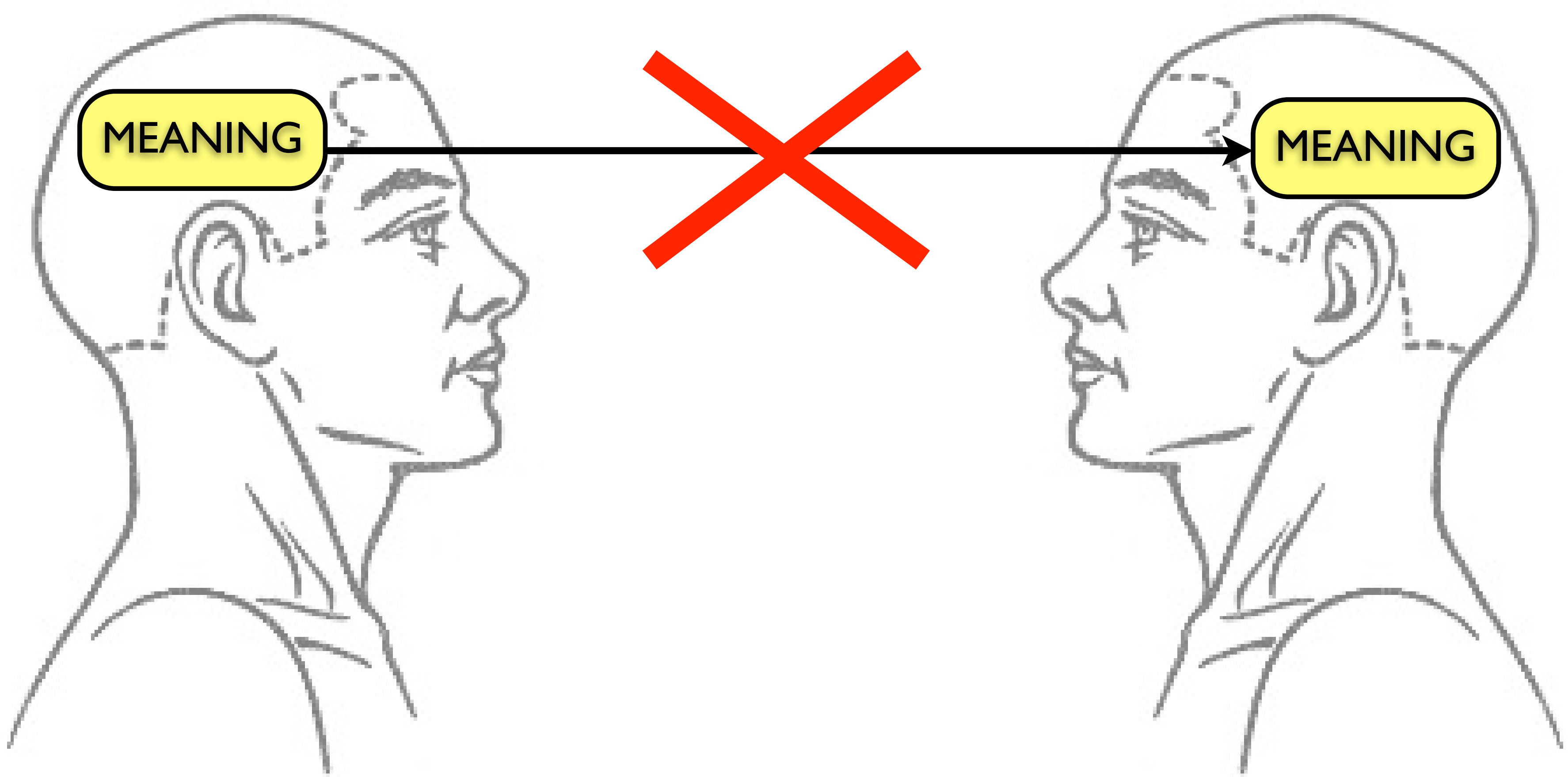
- It can be helpful to regard communication as the transfer of information from one individual's mind to another.



Direct Meaning Transfer

- But we are not telepathic: information cannot be transmitted directly between minds.

    - If we *could* transfer meanings, then why would we need signals at all?



Indirect Meaning Transfer

- The meaning must therefore be transferred *indirectly*.
- The speaker produces some behaviour which:
  - tells the hearer that they are trying to communicate (communicative intention);
  - and enables the hearer to recover the information or meaning (informative intention).
- This is what signals are for.

Inference of Meaning

- The speaker's signal provides evidence about the meaning they want to convey.
- The hearer interprets the speaker's signal to work out the meaning they think the speaker intended to convey.
- How do they work it out?

Communication as a Code

- If agents have an efficient code to translate meanings into signals and vice versa, then communication is (relatively) trivial.

	s_1	s_2	s_3
m_1	1	0	0
m_2	0	1	0
m_3	0	0	1

$$m_1 \longleftrightarrow s_1$$

$$m_2 \longleftrightarrow s_2$$

$$m_3 \longleftrightarrow s_3$$

Code Problems

- But this requires that agents:
  - have the same meanings;
  - have the same (or at least compatible) signal-meaning mappings.
- How does this happen?
  - We’re going to look at simulating the acquisition of signal-meaning mappings.
  - (If you’re interested there is other work that models how agents can create their own meanings.)

Fast Mapping

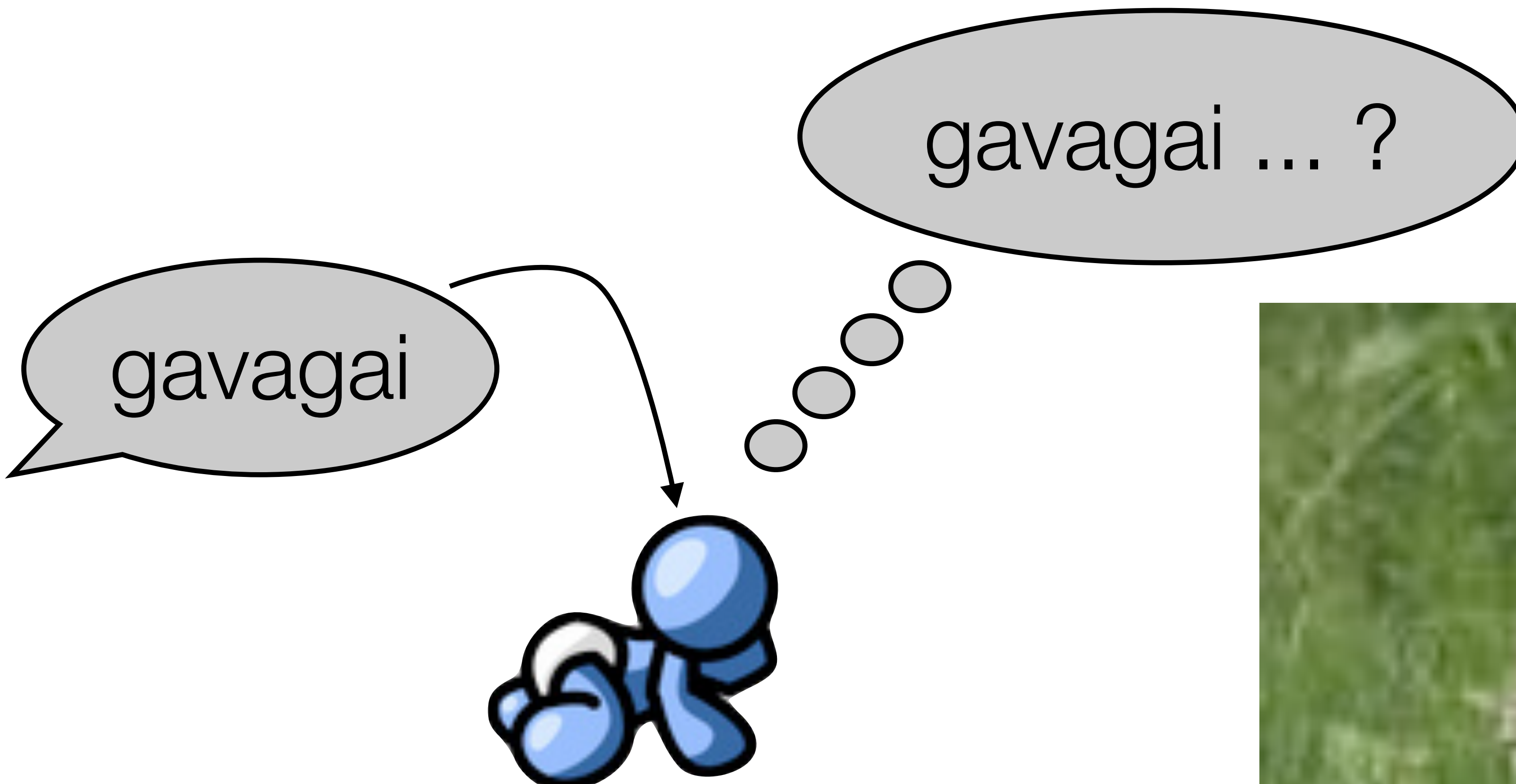
- Children can approximate a word’s meaning after a single exposure, through *fast mapping* (Carey and Bartlett 1978).

- Widespread assumption that fast mapping enables acquisition of large vocabularies (we learn ~ 60,000 word meanings by age 18).

- **But** shouldn’t it be very difficult to accurately infer the meaning of an unfamiliar word the first time you hear it?

Quine's Problem

- How does a learner work out the meaning of an unfamiliar signal?



- What does "gavagai" mean?

Indeterminacy of Translation

- rabbit
- grass
- some part(s) of the rabbit, or of the grass
- some property of some part of the rabbit (the colour of its ear)
- something the rabbit makes you think of (I'm hungry, fluffiness)
- something based on superstition (it will rain later)
- something even weirder (rabbits, but only till Scotland win the World Cup, then crows)

Indeterminacy of Meaning

- Quine showed that:

  1. there are infinitely many possible meanings for “gavagai” consistent with this particular usage episode.

  2. there are infinitely many possible meanings consistent with *any* possible sequence of usage episodes.

- But despite this, children *do* learn the meanings of words.

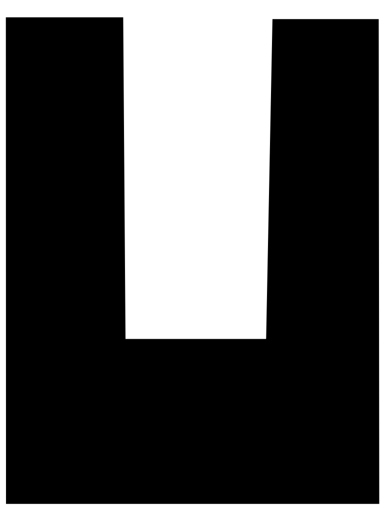
  - How?

Heuristics for the Inference of Meaning

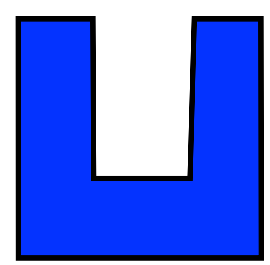
- Various strategies have been suggested for how children eliminate spurious meanings:
  - Behavioural cues to identify the attentional focus of the speaker (Baldwin 1991, Tomasello & Farrar 1986)
  - Expectations about what things are likely to be referents (Macnamara 1972, Landau et al 1988)
  - Expectations about words (Markman & Wachtel 1988, Doherty 2004)
  - Syntactic context (Gillette et al 1999)

This is a dax.

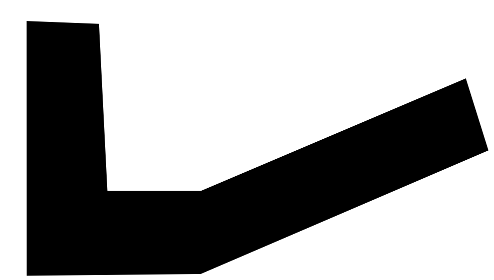
is this a dax?



is this?



is this?



Landau et al 1988

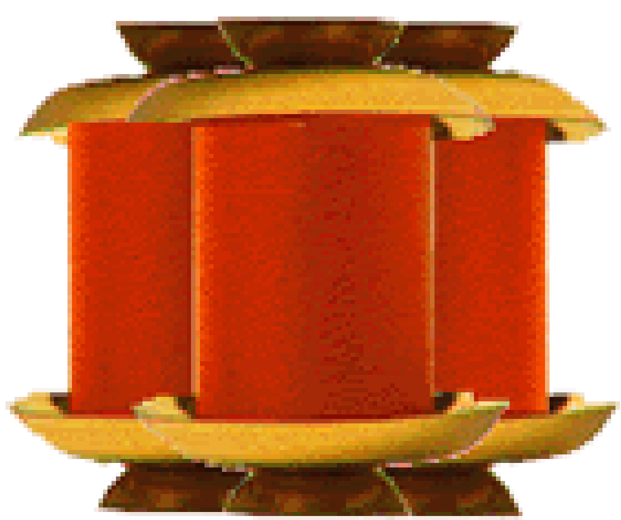
Reducing referential uncertainty

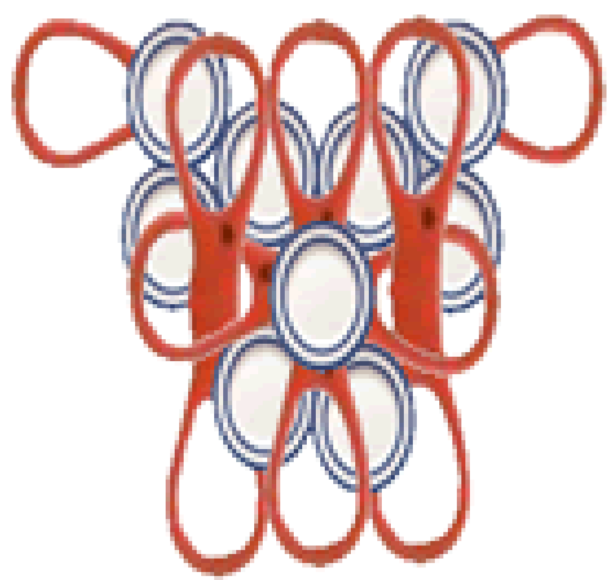
- These heuristics help to reduce referential uncertainty by eliminating spurious candidate meanings.
- Fast mapping requires the elimination of *all* uncertainty.
    - This is probably very hard work, and probably requires a very helpful learning context
- So what can you do if you are always left with two or more possible meanings for a word?
    - use information you get from hearing the word in different contexts.
- This is *cross-situational learning* (Siskind 1996, Blythe et al, 2010, Smith et al, 2011).

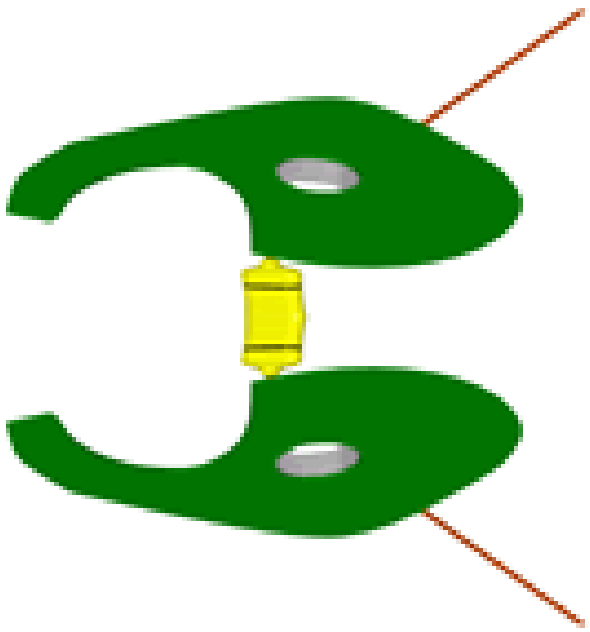
Cross-situational learning

- What does “quidector” refer to?

quidector

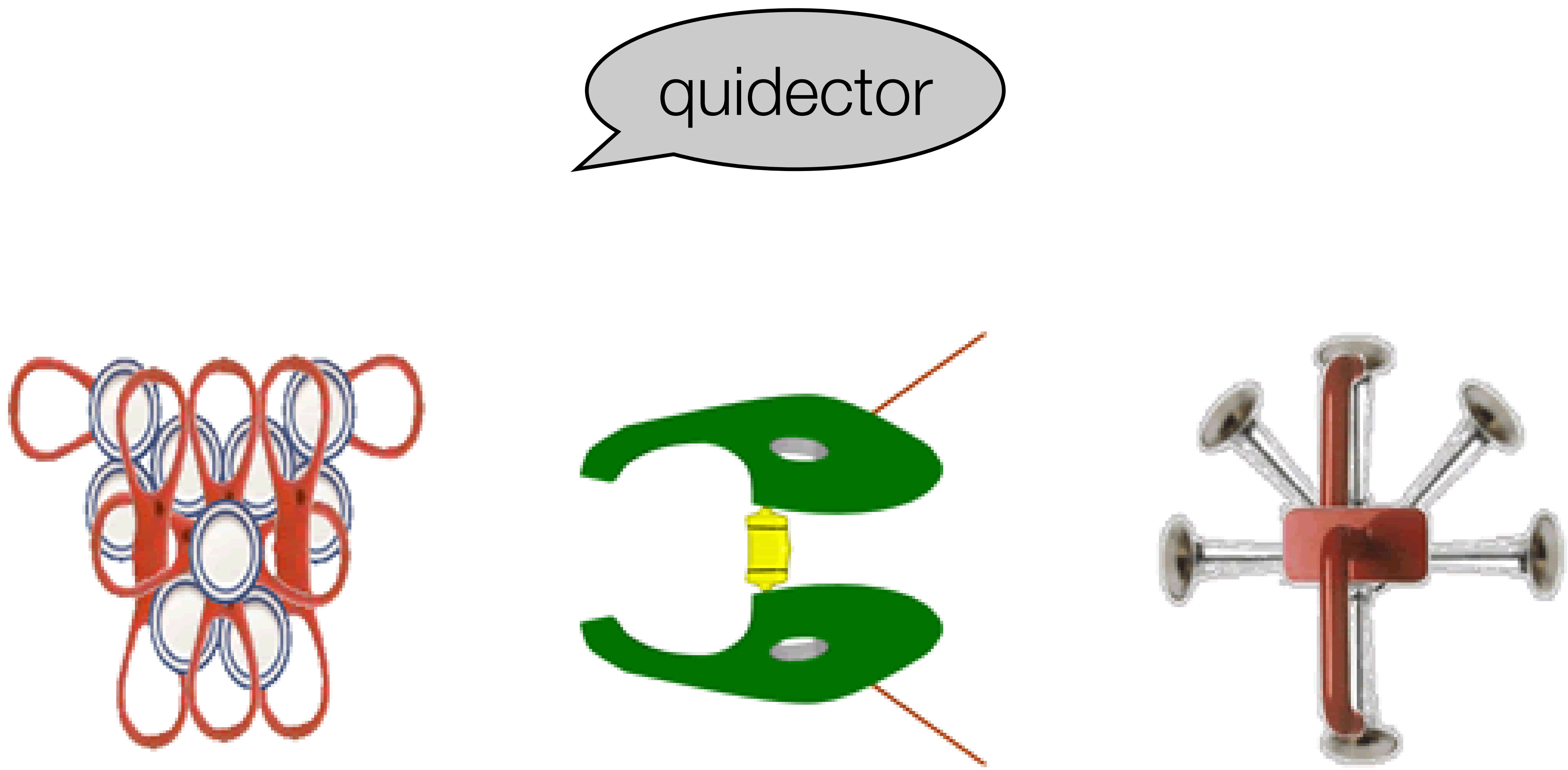






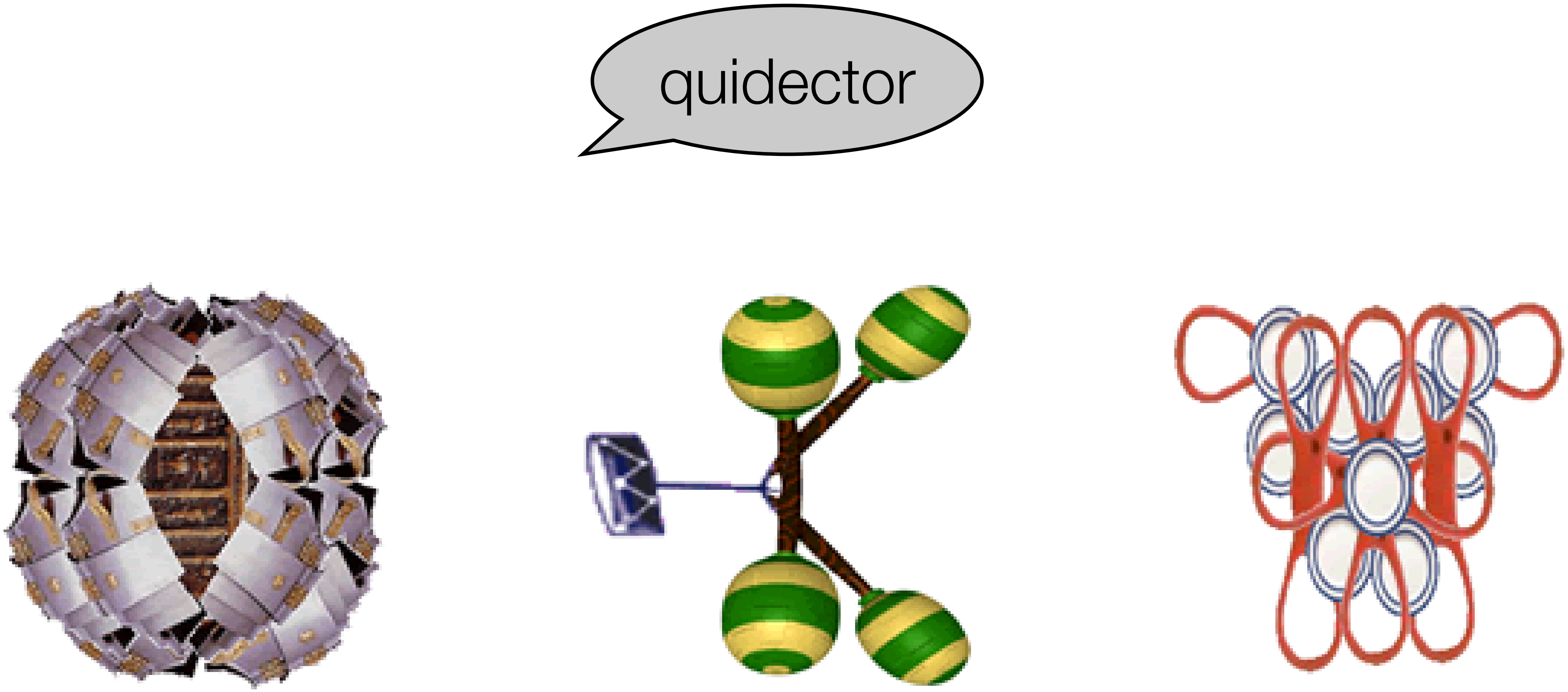
Cross-situational learning

- Now what do you think “quidector” refers to?



Cross-situational learning

- And now?

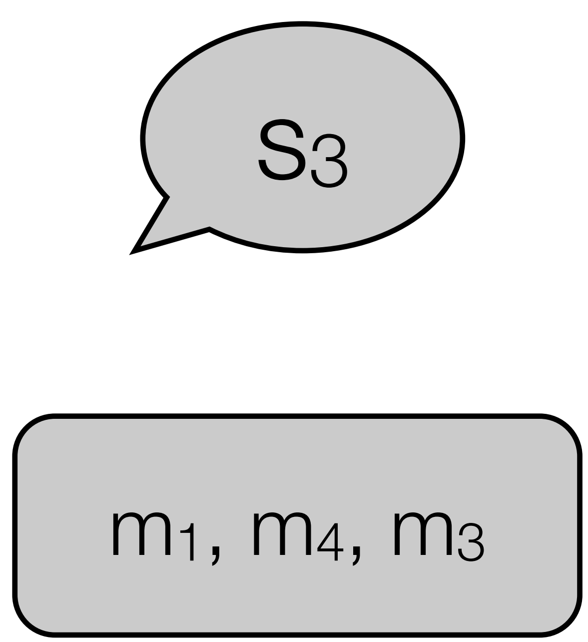


Context

- Cross-situational learning is based on the co-occurrence of signals and meanings across multiple learning (or communicative) episodes.
- During each episode, the *context* provides a set of candidate meanings.
- Each of these meanings is associated with the signal.
- The intersection of the various sets of candidate meanings at each exposure will yield the ‘true’ meaning.

Cross-situational learning data

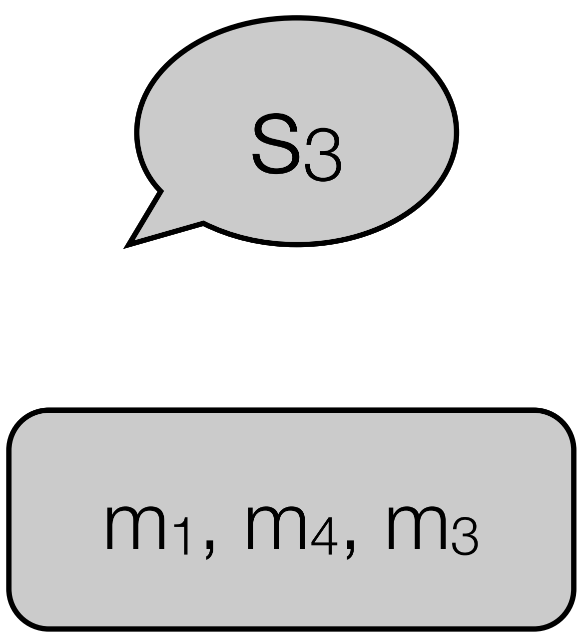
- Instead of a meaning-signal pair, we assume that the learner:
  - hears a signal;
  - and the context provides a set of meanings.



	s_1	s_2	s_3	s_4	s_5
m_1	0	0	0	0	0
m_2	0	0	0	0	0
m_3	0	0	0	0	0
m_4	0	0	0	0	0
m_5	0	0	0	0	0

Cross-situational learning data

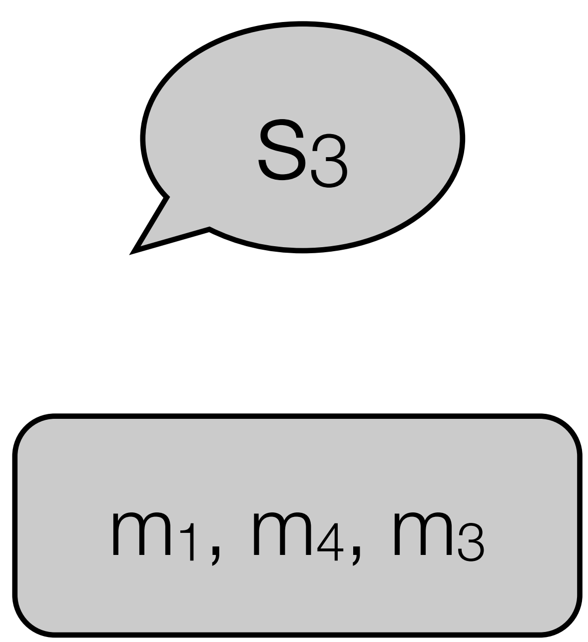
- Instead of a meaning-signal pair, we assume that the learner:
    - hears a signal;
    - and the context provides a set of meanings.



	s_1	s_2	$\mathbf{s_3}$	s_4	s_5
$\mathbf{m_1}$	0	0	0	0	0
m_2	0	0	0	0	0
$\mathbf{m_3}$	0	0	0	0	0
$\mathbf{m_4}$	0	0	0	0	0
m_5	0	0	0	0	0

Cross-situational learning data

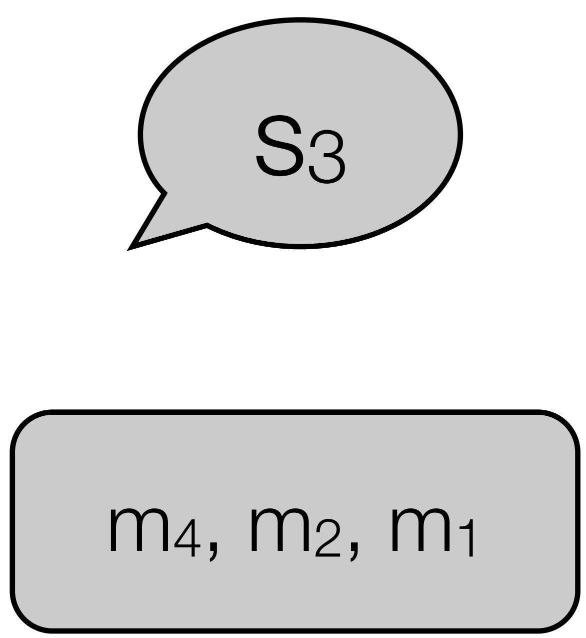
- Instead of a meaning-signal pair, we assume that the learner:
    - hears a signal;
    - and the context provides a set of meanings.



	s_1	s_2	s_3	s_4	s_5
m_1	0	0	**1**	0	0
m_2	0	0	0	0	0
m_3	0	0	**1**	0	0
m_4	0	0	**1**	0	0
m_5	0	0	0	0	0

Cross-situational learning data

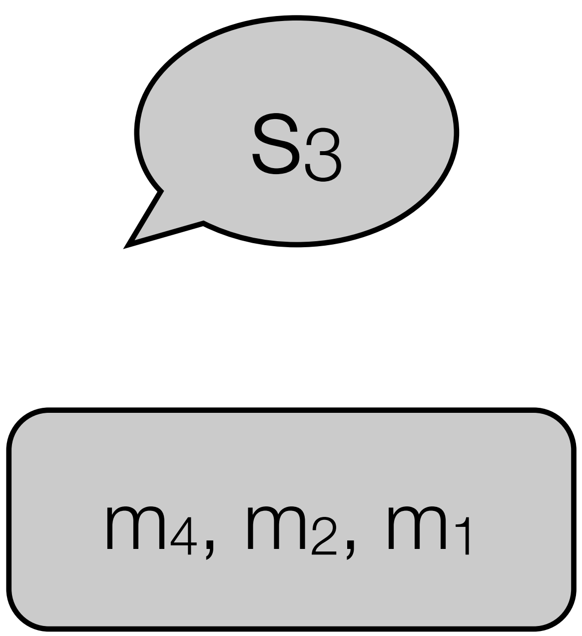
- The next episode has a different context, which provides a different set of candidate meanings.



	s_1	s_2	$\mathbf{s_3}$	s_4	s_5
$\mathbf{m_1}$	0	0	1	0	0
$\mathbf{m_2}$	0	0	0	0	0
m_3	0	0	1	0	0
$\mathbf{m_4}$	0	0	1	0	0
m_5	0	0	0	0	0

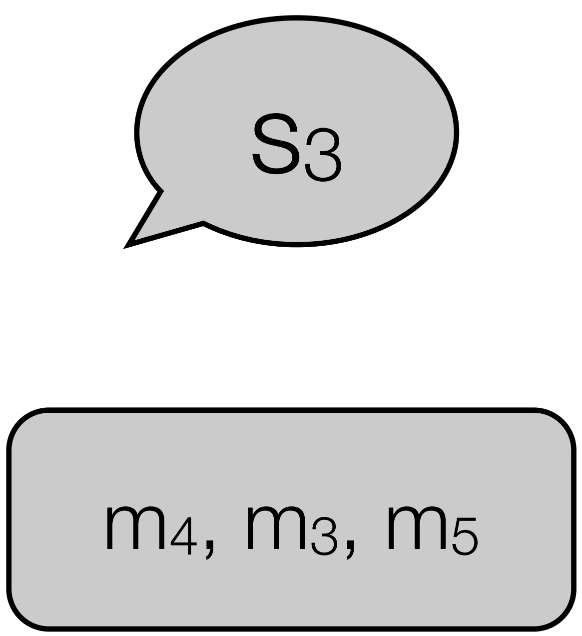
Cross-situational learning data

- The next episode has a different context, which provides a different set of candidate meanings.



	s_1	s_2	s_3	s_4	s_5
m_1	0	0	**2**	0	0
m_2	0	0	**1**	0	0
m_3	0	0	1	0	0
m_4	0	0	**2**	0	0
m_5	0	0	0	0	0

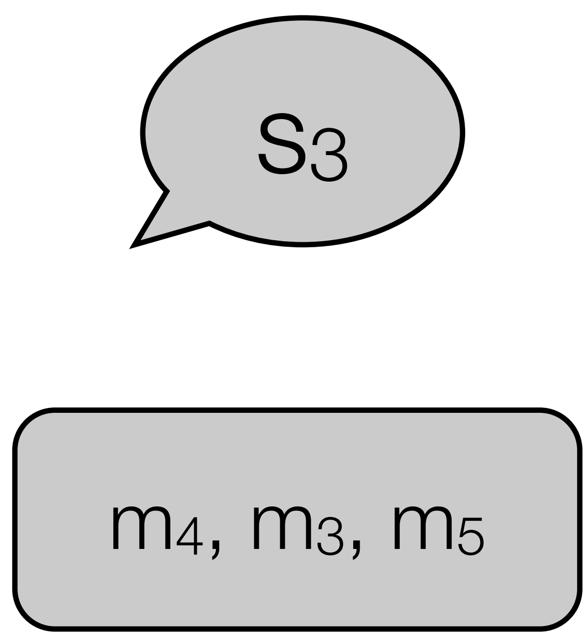
Cross-situational learning data



	s_1	s_2	$\mathbf{s_3}$	s_4	s_5
m_1	0	0	2	0	0
m_2	0	0	1	0	0
$\mathbf{m_3}$	0	0	1	0	0
$\mathbf{m_4}$	0	0	2	0	0
$\mathbf{m_5}$	0	0	0	0	0

Cross-situational learning data

- Eventually, the cross-situational information reveals the true meaning.



	s_1	s_2	s_3	s_4	s_5
m_1	0	0	2	0	0
m_2	0	0	1	0	0
m_3	0	0	**2**	0	0
m_4	0	0	**3**	0	0
m_5	0	0	**1**	0	0

Learning rules and cross-situational learning

- Let's think of cross-situational learning in terms of Smith (2002)'s characterisation of learning rules.

- Previously, there was always:
    - 1 cell to which α applies
    - s-1 cells to which β applies
    - m-1 cells to which γ applies
    - δ applies to all the rest (s-1)*(m-1)

	s_1	s_2	s_3	s_4	s_5
m_1	**α**	**β**	**β**	**β**	**β**
m_2	**γ**	δ	δ	δ	δ
m_3	**γ**	δ	δ	δ	δ
m_4	**γ**	δ	δ	δ	δ
m_5	**γ**	δ	δ	δ	δ

Learning rules and cross-situational learning

- In cross-situational learning, there is not one but C (the size of the context) meanings active at the same time as the signal.
- This increases the number of cells to which α and β apply, and decreases the number to which γ and δ apply.
  - α applies to C cells
  - β applies to (s-1)C cells
  - γ applies to m-C cells
  - δ applies to the rest (s-1)*(m-C)

	s_1	s_2	s_3	s_4	s_5
m_1	**α**	**β**	**β**	**β**	**β**
m_2	**α**	**β**	**β**	**β**	**β**
m_3	**α**	**β**	**β**	**β**	**β**
m_4	**γ**	δ	δ	δ	δ
m_5	**γ**	δ	δ	δ	δ

Slow Learning?

- The time taken to learn a lexicon through cross-situational learning depends on:
  - the size of the context at each learning episode.
  - the number of meanings in the lexicon
- Cross-situational learning is clearly slower than immediate fast mapping would be.
  - But how much slower?

Testing cross-situational learning

- Mathematical studes show that cross-situational learning can account for learning large lexicons, without the need for very strong heuristics: there's no link between learning individual words rapidly and being able to acquire a large lexicon (Blythe et al, 2010).

- Experimental studies show that humans are capable of cross-situational learning (Akhtar & Montague 1999, Gilette et al. 1999, Houston-Price et al 2003, Yu & Smith 2007, Smith et al 2009, Smith et al 2011).

  - but that the rigour with which we use cross-situational learning depends on the difficulty of the task - how large the size of the context is compared to the size of the lexicon, or how the data is presented (Smith et al 2011).

Reading for this lecture

- Siskind, J. M. (1996) A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition* 61:1-38.