Beyond “feeled dat”

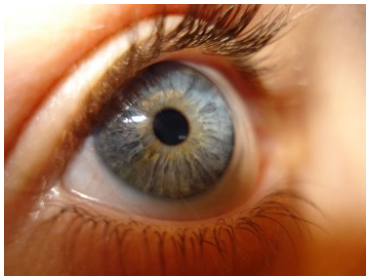
6.S077 Recitation 9

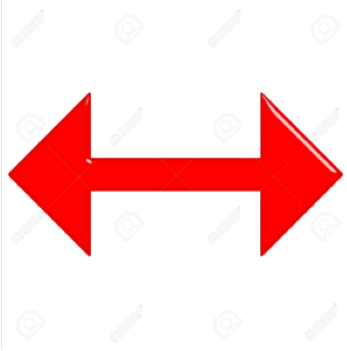
Language & words:
The Indeterminacy Problem

Possible Hypotheses

- **Rabbit (whole object)**
- **Animal (superordinate)**
- **Flopsie (individual)**
- **Furry (property)**
- **Ear (part)**
- **Walk by (activity)**
- **Undetached rabbit parts .....**

A picture is worth a thousand words – and that's the problem!





David Hume (1711-1776)

Though a particular colour, taste and smell, are qualities all united together in ***this apple,*** it is easy to perceive they are not the same, but are at least distinguishable from each other.

Learning by “ostension”:
“We show them the thing and tell them the name that stands for it”

Constraints Guide Learning

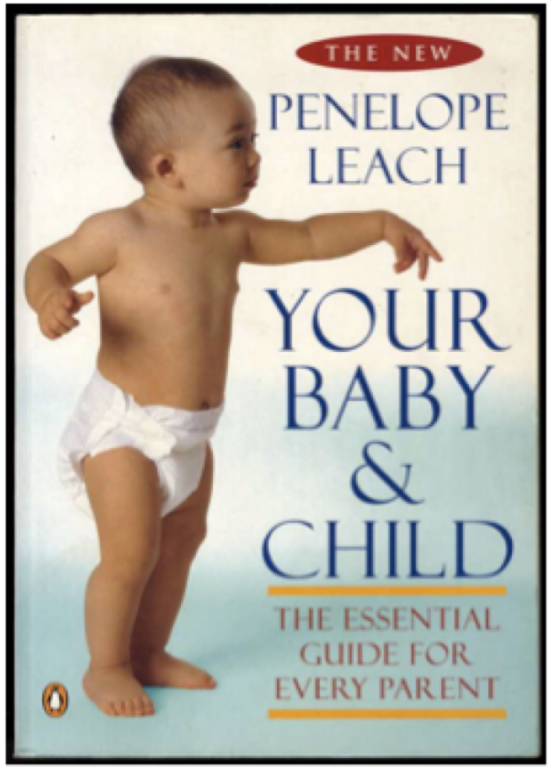
(Brown 1957, Rosch, Markman, Clark, Baldwin, Gleitman, Landau, ...)

- Syntactic Cues

But also:

- "Basic Level" Constraint
- Whole Object Constraint
- Mutual Exclusivity
- Eye gaze-referent
- Shape bias
- Theory of mind

How do children learn words?



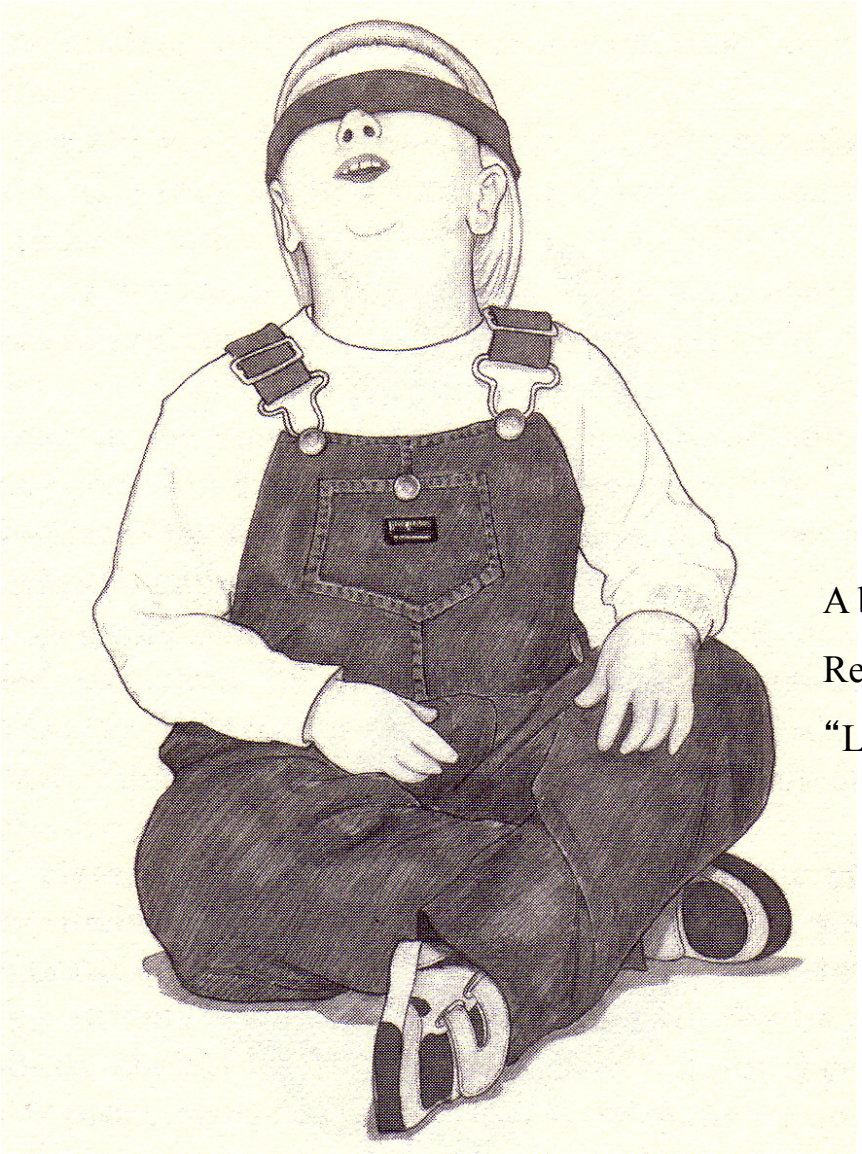
A baby hears a word like “shoes”, for example, over and over again in daily life as the one constant sound in a large variety of statements.

Laboratory shoe world

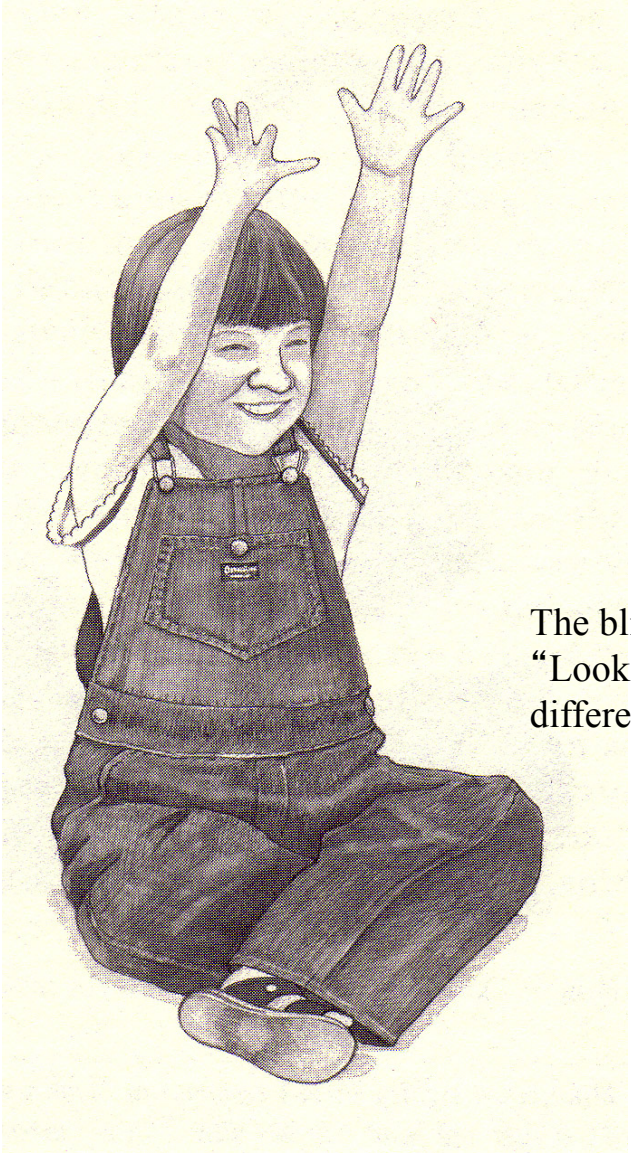
The real shoe world

What regularities are used to learn words?

Is this doll hard to see?



A blindfolded sighted child
Responds to the command
“Look up!”



The blind child's response to "Look up!" is the same and different.

Two kinds of evidence – how syntax or situation can help ("bootstrapping")

1. Within arm's reach [**the world**]
2. Exploratory and perceptual [**the structures**]:

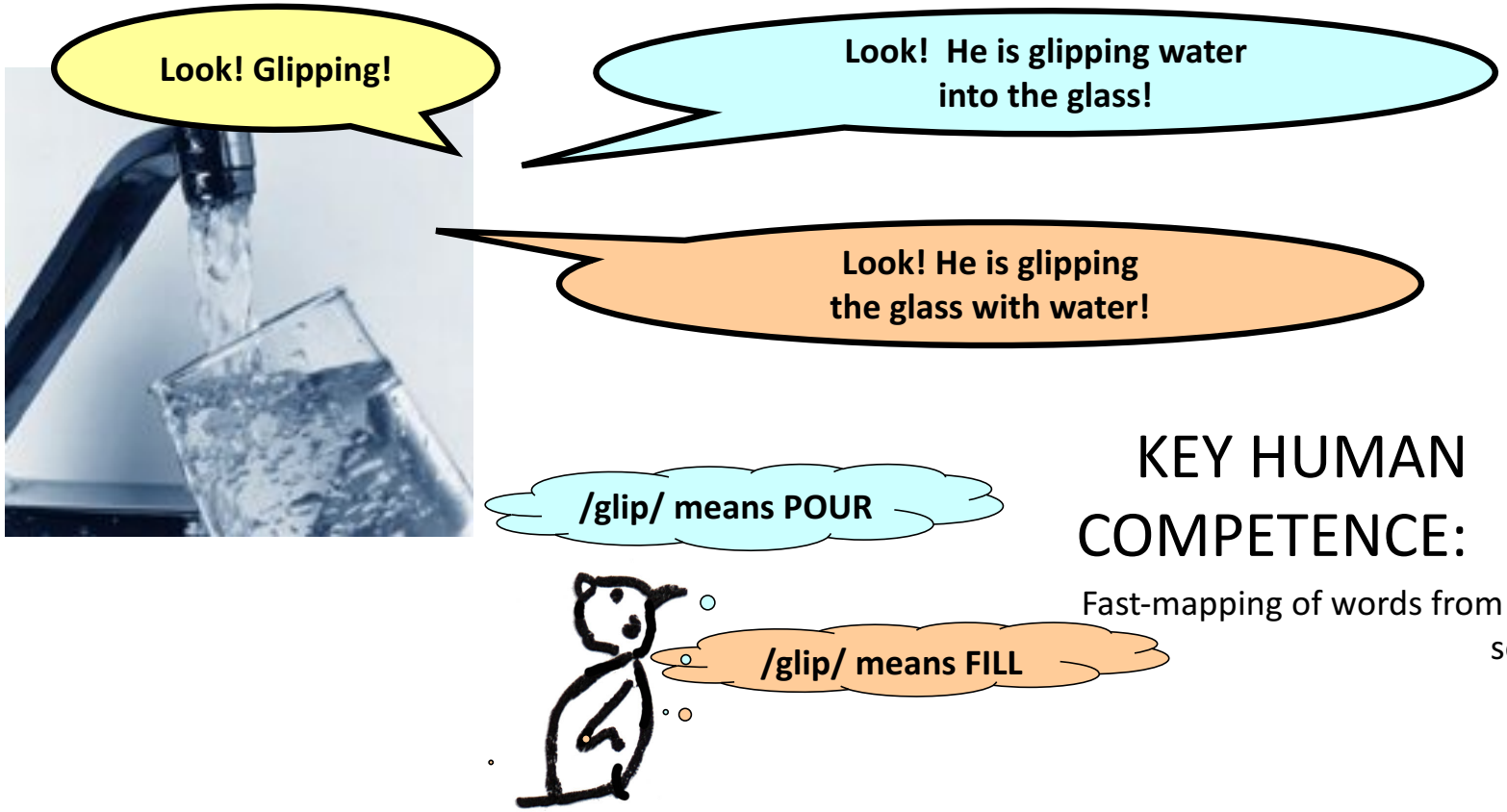
Can you see the cheese? [NP complement] because you can observe *things*]

Let's see if there's cheese in the refrigerator [clausal complement because you can observe events]

Let's see where the cheese is [because these are spatial perceptions]

We are looking at/*seeing/ the cheese [acts and states]

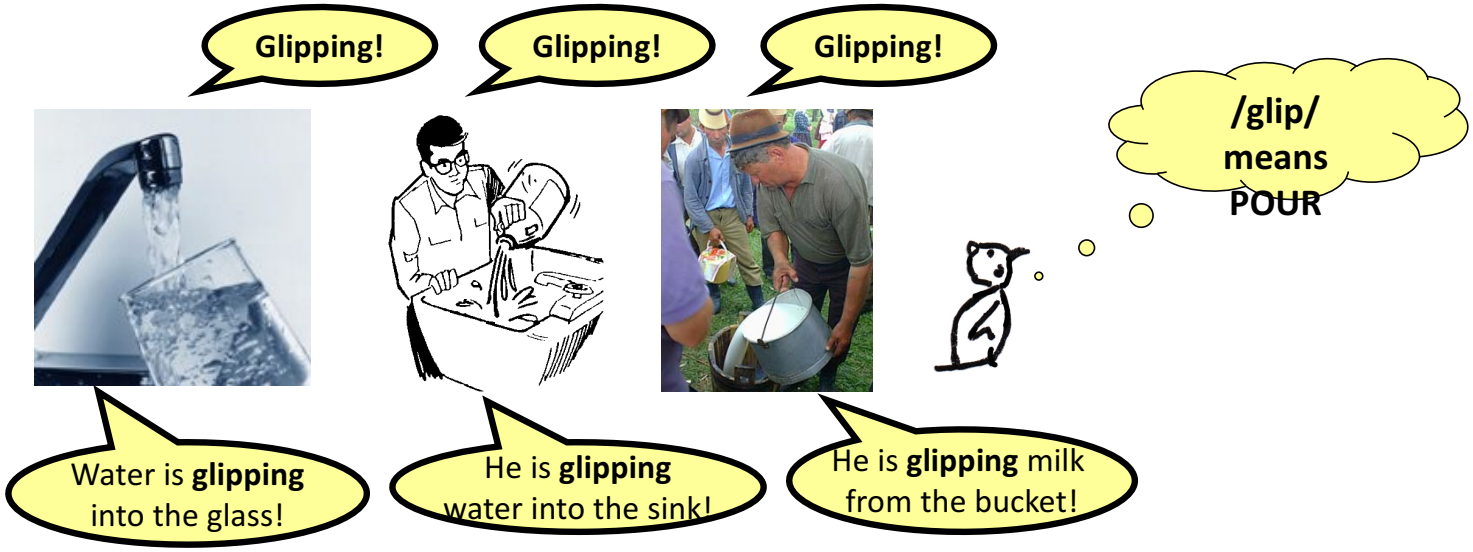
Syntactic Bootstrapping



KEY HUMAN COMPETENCE:

Fast-mapping of words from syntax & semantics

Semantic Bootstrapping



Syntactic Bootstrapping: Distribution of *syntactic frames* disambiguates

(Gleitman 1990, Naigles 1990, Fisher et al 1994, …)

Two Bootstrapping Proposals

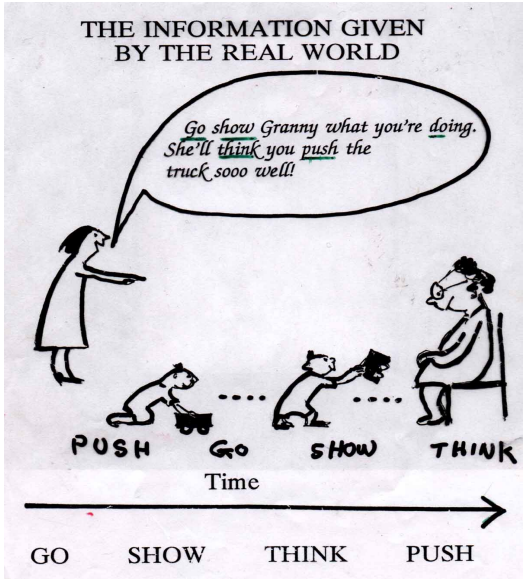
- Children use syntactic cues to verb meaning (Gleitman 1990)
- Children use (verb) meaning to figure out how its arguments are realized in the syntax of the language (Pinker 1989)

Semantic Bootstrapping

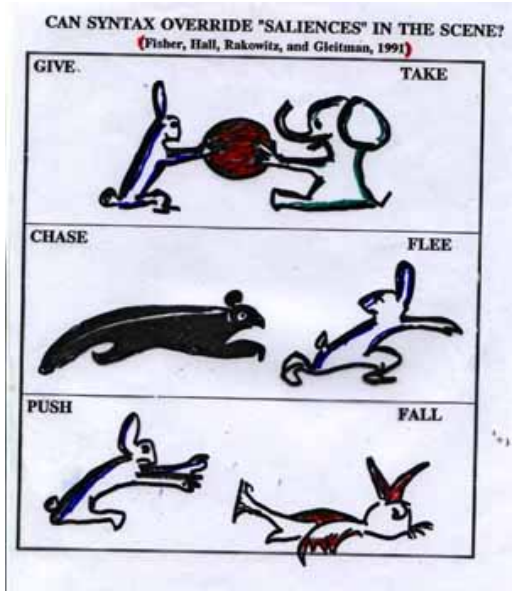
Semantic Bootstrapping involves the pairing of a situational context with some syntactic pattern

- Kids learn syntax by first learning the semantic argument structure of the verb.
  - SWIM = one participant (the “swimmer”)
  - EAT = two participants (“eater”, “eatee”)
  - TAKE = two/three participants (“taker”, “takee”, and “person taken from”…)

But there are ambiguities



Temporal ambiguity



Situation ambiguity

Mental unobservable!

... more than just real-world observation...

Syntactic Bootstrapping

(Landau and Gleitman 1986, Naigles 1990)

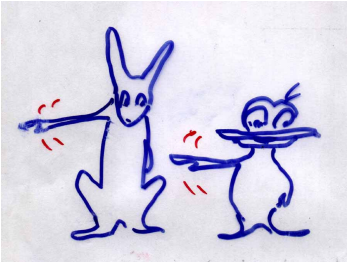
Syntactic frames provide *evidence* for meaning:

/X and Y are gorping!/

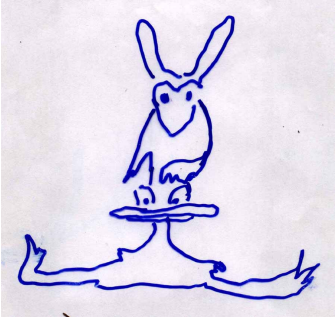
/X is gorping Y!/

Find gorping!

H_1: arm wheel



H_2: cause to squat



Language and *n*-grams (RR #2)

N-grams: Unigrams

- Given a corpus of text, the *n*-grams are the sequences of *n* consecutive words that are in the corpus
- example (*12 word sentence*)
  - the cat that sat on the sofa also sat on the mat
- *n*=1 (8 ***unigrams***)
  - the 3
  - sat 2
  - on 2
  - cat 1
  - that 1
  - sofa 1
  - also 1
  - mat 1

N-grams: Bigrams

- **example** (12 word sentence)
  - the cat that sat on the sofa also sat on the mat

    2 words
- N=2 (8 ***bigrams***)
  - sat on 2
  - on the 2
  - the cat 1
  - cat that 1
  - that sat 1
  - the sofa 1
  - sofa also 1
  - also sat 1
  - the mat 1

N-grams: Trigrams

- **example** (12 word sentence)
  - the cat that sat on the sofa also sat on the mat

    3 words
- N=3 (9 **trigrams**)
  - *most language models stop here, some stop at quadrigrams*
    - too many n-grams
    - low frequencies
  - sat on the 2
  - the cat that 1
  - cat that sat 1
  - that sat on 1
  - on the sofa 1
  - the sofa also 1
  - sofa also sat 1
  - also sat on 1
  - on the mat 1

‘Modeling’ sentences

It was a bright cold day in April and the clock was striking…

Step 1, use chain rule

$$
\begin{aligned}
P(S) &= P(w_1, \ldots, w_n) \\
&= P(w_1)P(w_2|w_1)P(w_3|w_1, w_2) \ldots P(w_n|w_1, \ldots, w_{n-1}) \\
&= \prod_{i=1}^{n} P(w_i|w_1, \ldots, w_{i-1})
\end{aligned}
$$

$$
\begin{aligned}
P(S) = {} & P(it) \times P(was|it) \times P(a|it, was) \times P(bright|it, was, a) \times \ldots \times \\
& P(April|it, was, a, bright, cold, day, in)
\end{aligned}
$$

Step 2: use Markov assumption to limit 'history' length (context)

Make the Markov assumption, either order 1 (1 word history), or order 2 (2 word history):

$$P(w_i|w_1,\ldots,w_{i-1}) \approx P(w_i|w_{i-1})$$
$$P(w_i|w_1,\ldots,w_{i-1}) \approx P(w_i|w_{i-2},w_{i-1})$$

trigram model:

$$P(S) \approx P(it) \times P(was|it) \times P(a|it,was) \times P(bright|was,a) \times \ldots \times P(April|day,in)$$

Use *counts* as maximum likelihood estimates of ‘true’ probabilities

$$P_{MLE}(w_i|w_{i-1}) = \frac{C(w_i, w_{i-1})}{C(w_{i-1})}$$

Let’s try this out on the corpus G. Orwell, “1984”
Training set: main text; test set: Appendix, “Principles of Newspeak” – 1st sentence is:

<s>*A good deal of the literature of the past was, indeed, already being transformed in this way* </s>

115,212 words
8,635 types
3,928 unique words (‘hapax legomena’)
49,524 bigrams
37,365 bigrams w/ frequency of 1

Where do estimates come from?

Let $S = w_1, w_2, \ldots, w_n$ be a word sequence.
Given a training corpus, an intuitive estimate of the probability of the sequence P(S), is the relative frequency of the string $w_1, w_2, \ldots, w_n$ in the corpus, the *maximum likelihood estimate* (MLE):

$$P_{MLE} = \frac{C(w_1, \ldots, w_n)}{N}$$

where $C(w_1, \ldots, w_n)$ is the frequency or count of the string $w_1, \ldots, w_n$ in the corpus and N is the total number of strings of length n.