Computational Cognitive Science

Lecture 13: Word Learning

Stella Frank
stella.frank@ed.ac.uk

29 October 2019

Word Learning: Frank et al. (2009)

Evidence for word learning

Word learning could use two types of evidence:

Statistical information about word-object co-occurrences

Pragmatic intention: what is the speaker trying to tell me?

Evidence for word learning

Word learning could use two types of evidence:

Statistical information about word-object co-occurrences

Pragmatic intention: what is the speaker trying to tell me?

What is developmentally *plausible* about each type of evidence?
What is developmentally *implausible* about each type of evidence?

Evidence for word learning

Word learning could use two types of evidence:

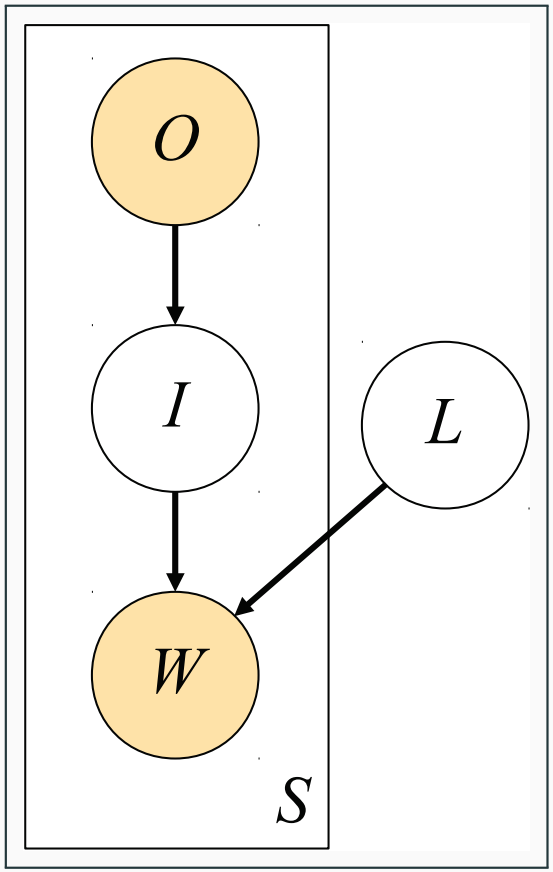
Statistical information about word-object co-occurrences

Pragmatic intention: what is the speaker trying to tell me?

What is developmentally *plausible* about each type of evidence?
What is developmentally *implausible* about each type of evidence?

Goal of Frank et al. (2009) model: combine both types of evidence into one Bayesian model

M. Frank et al. (2009): Using Speakers Referential Intentions to Model Early Cross-Situational Word Learning

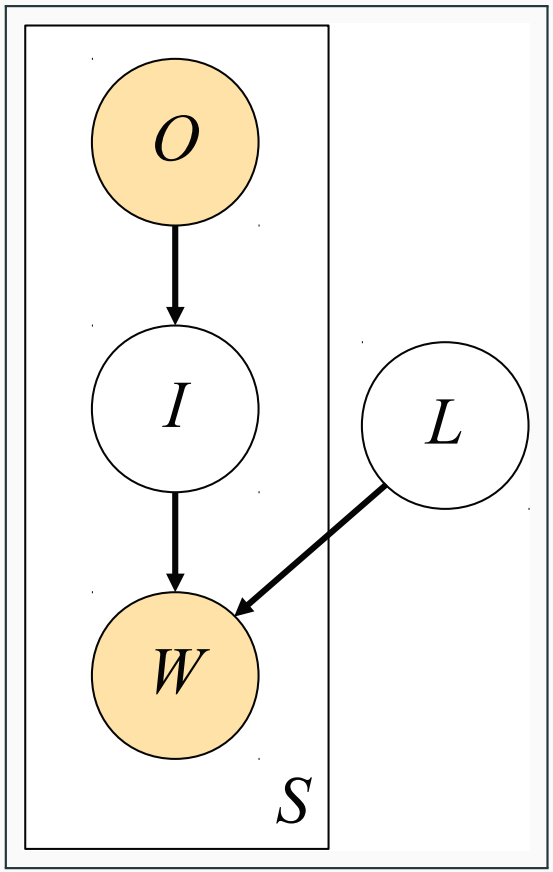


Frank et al. (2009)

Graphical model notation:

- empty/white-background circle: hidden/latent random variable
- shaded/colored circle: observed random variable
- arrow: conditional dependence
- plate: replicated S times.

M. Frank et al. (2009): Using Speakers Referential Intentions to Model Early Cross-Situational Word Learning

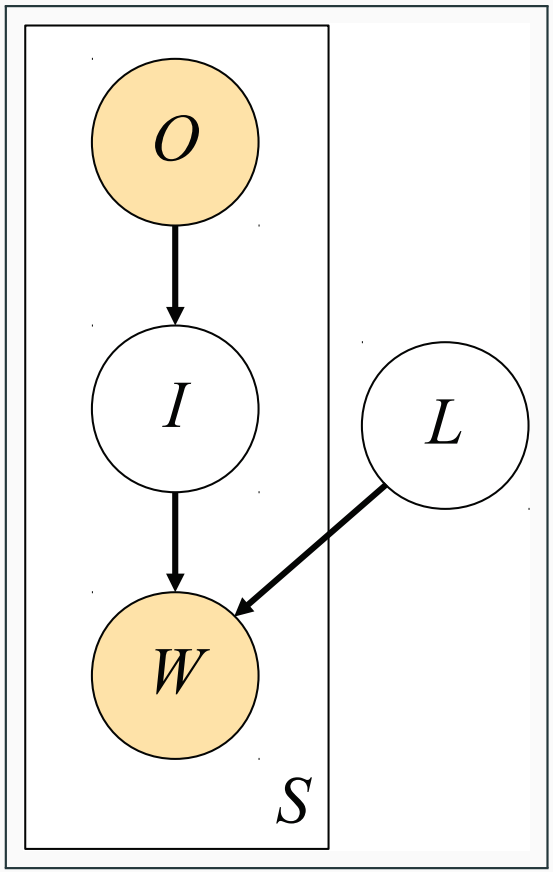


Frank et al. (2009)

For each situation (utterance) $s \in S$:

- Objects O are present and observable.
- The speaker chooses a set of intended referents $I \subseteq O$, not visible to the learner (a hidden/latent variable).
- The speaker chooses a set of words $W \in L \cup C$
  - Some of these words are used referentially, to refer to referents in I, using words from the lexicon L
  - Others are not: these can be words in L or other words from the corpus C

M. Frank et al. (2009): Using Speakers Referential Intentions to Model Early Cross-Situational Word Learning



Frank et al. (2009)

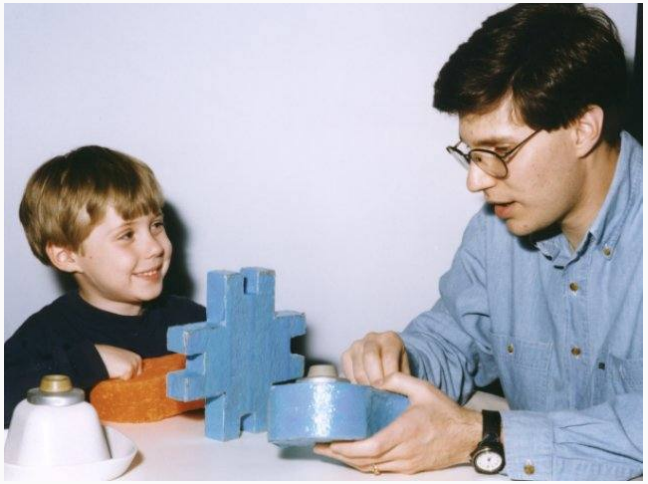
For each situation (utterance) $s \in S$:

- Objects O are present and observable.
- The speaker chooses a set of intended referents $I \subseteq O$, not visible to the learner (a hidden/latent variable).
- The speaker chooses a set of words $W \in L \cup C$
  - Some of these words are used referentially, to refer to referents in I, using words from the lexicon L
  - Others are not: these can be words in L or other words from the corpus C

Goal is to infer values (or distributions over values) for the latent variables, here L and I.

Why model intentions?

Children rely on *social context* to learn words. They infer other peoples' intentions based on:



- eye gaze;
- body position;
- pragmatics/saliency.

Evidence for this comes from:

- tracking of others' gaze at six months;
- learning new words using gaze at 18 months.

Note that Frank et al. (2009) model doesn't use these extra-linguistic cues. Two models that do are:

Yu & Ballard (2007) A unified model of early word learning: Integrating statistical and social cues. Neurocomputing

Frank et al. (2008) A Bayesian Framework for Cross-Situational Word-Learning. NIPS

Bayesian Model (Frank et al., 2009)

Posterior:

$$P(L|D) \propto P(D|L)P(L)$$

Prior:

$$P(L) \propto e^{-\alpha|L|}$$

Likelihood:

$$\begin{aligned}
P(D|L) &= \prod_{s \in D} P(O_s, W_s|L) \\
&= \prod_{s \in D} \sum_{I_s \subseteq O_s} P(O_s, I_s, W_s|L) \\
&= \prod_{s \in D} \sum_{I_s \subseteq O_s} P(O_s)P(I_s|O_s)P(W_s|I_s, L) \\
&\propto \prod_{s \in D} \sum_{I_s \subseteq O_s} P(I_s|O_s)P(W_s|I_s, L)
\end{aligned}$$

Generative Model

Likelihood: $P(D|L) \propto \prod_{s \in D} \sum_{I_s \subseteq O_s} P(I_s|O_s) P(W_s|I_s, L)$

Generate intentions from objects: uniform distribution:

$$P(I_s|O_s) \propto 1$$

Generate words from intentions and lexicon $P(W_s|I_s, L)$, where words are independent from each other. For each word w in W_s:

- choose referential ($p = \gamma$) or non-referential ($p = 1 - \gamma$):

$$P(W_s|I_s, L) = \prod_{w \in W_s} \left[\gamma \sum_{o \in I_s} \frac{1}{|I_s|} P_R(w|o, L) + (1 - \gamma) P_{NR}(w|L) \right]$$

- $P_R(w|o, L)$: choose uniformly from lexical items that refer to correct object;
- $P_{NR}(w|L)$: choose quasi-uniformly from all words in corpus: probability 1 if word is not in lexicon, κ if word is in lexicon.

Inference

Search for best (MAP, *maximum a posteriori*) solution.

Solution: Lexicon only, marginalising over intentions

Possible difficulties:

Inference

Search for best (MAP, *maximum a posteriori*) solution.

Solution: Lexicon only, marginalising over intentions

Possible difficulties:

- Search space is very large
- Posterior is also very un-smooth, making it:
- Easy to get stuck in local maxima
- Hard to find the right direction to get out

See the Supplementary Materials/Technical Appendix for their strategy.

Data

Word Learning: Infant's Input

`http://childes.talkbank.org/browser/index.php?url=Eng-NA/Rollins/me03.c`

Child is $\sim$ 9 months old.

Rollins, P. R., (2003) Caregivers' contingent comments to 9-month-old infants: Relationships with later language. Applied Psycholinguistics 24, 221–234

Frank et al. and Yu & Ballard use same two files, `me03` and `di06`. (Frank et al. has a typo: `me06` and `di03`)

Word Learning: Infant's Input

Two ten-minute videos annotated as $\sim$ 600 utterances;
$\sim$ 400 word types; $\sim$ 20 objects types.

- Transcript: *You want to do the rattle*
- Objects: `Face Mother Rattle Chair`
- Intentions: `Rattle`

Word Learning: Infant's Input

Two ten-minute videos annotated as $\sim$ 600 utterances;
$\sim$ 400 word types; $\sim$ 20 objects types.

- Transcript: *You want to do the rattle*
- Objects: `Face Mother Rattle Chair`
- Intentions: `Rattle`

In Matlab: `corpus.mat`: everything is an array of ints

- Words: `[6,205,391,66,293,50,84]`
- Objects: `[4,3,19,10]`

Results

Evaluation

Evaluate model predictions against:

- gold-standard lexicon (word-object pairings);
- gold-standard intentions for each utterance (coded manually).

Compute *precision* (proportion of pairings that were correct) and *recall* (proportion of total correct pairings that were found).

F-score is the harmonic mean of precision and recall.

Compare against related models:

- simple statistics (co-occurrence frequency, conditional probability, mutual information);
- cross-situational model without intentions: IBM Machine Translation Model 1.

Results: Lexicon Accuracy

Model	Precision	Recall	F score
Association frequency	.06	.26	.10
Conditional probability (object|word)	.07	.21	.10
Conditional probability (word|object)	.07	.32	.11
Mutual information	.06	.47	.11
Translation model (object|word)	.07	.32	.12
Translation model (word|object)	.15	.38	.22
Intentional model	**.67**	**.47**	**.55**
Intentional model (one parameter)	.57	.38	.46

Frank et al. (2009)

Results: Lexicon Accuracy

Word	Object
bear	**bear**
bigbird	**bird**
bird	**duck**
birdie	**duck**
book	**book**
bottle	bear
bunnies	**bunny**
bunnyrabbit	**bunny**
hand	**hand**
hat	**hat**
hiphop	mirror
kittycat	**kitty**
lamb	**lamb**
laugh	cow
meow	baby
mhmm	hand
mirror	**mirror**
moocow	**cow**
oink	pig
on	ring
pig	**pig**
put	ring
ring	**ring**
sheep	**sheep**

Frank et al. (2009)

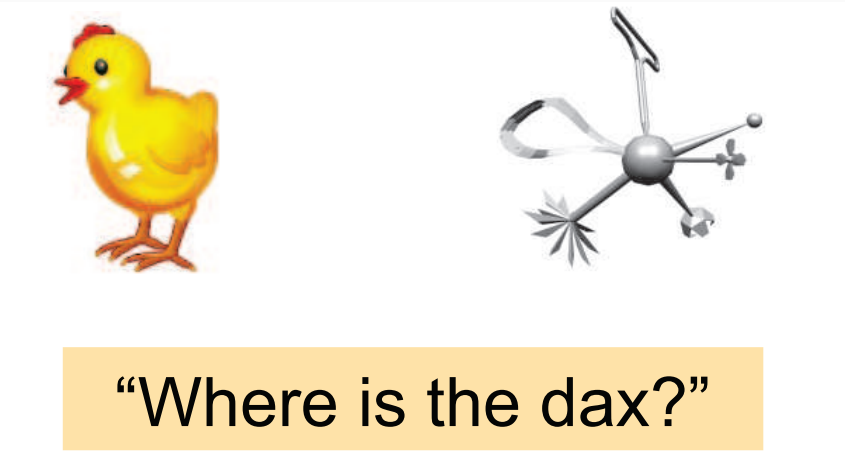
Results: Referent Accuracy

Model	Precision	Recall	F score
Association frequency	.27	**.81**	.40
Conditional probability (object|word)	.59	.36	.45
Conditional probability (word|object)	.32	.79	.46
Mutual information	.36	.37	.37
Translation model (object|word)	.57	.41	.48
Translation model (word|object)	.40	.57	.47
Intentional model	**.83**	.45	**.58**
Intentional model (one parameter)	.77	.36	.50

Frank et al. (2009)

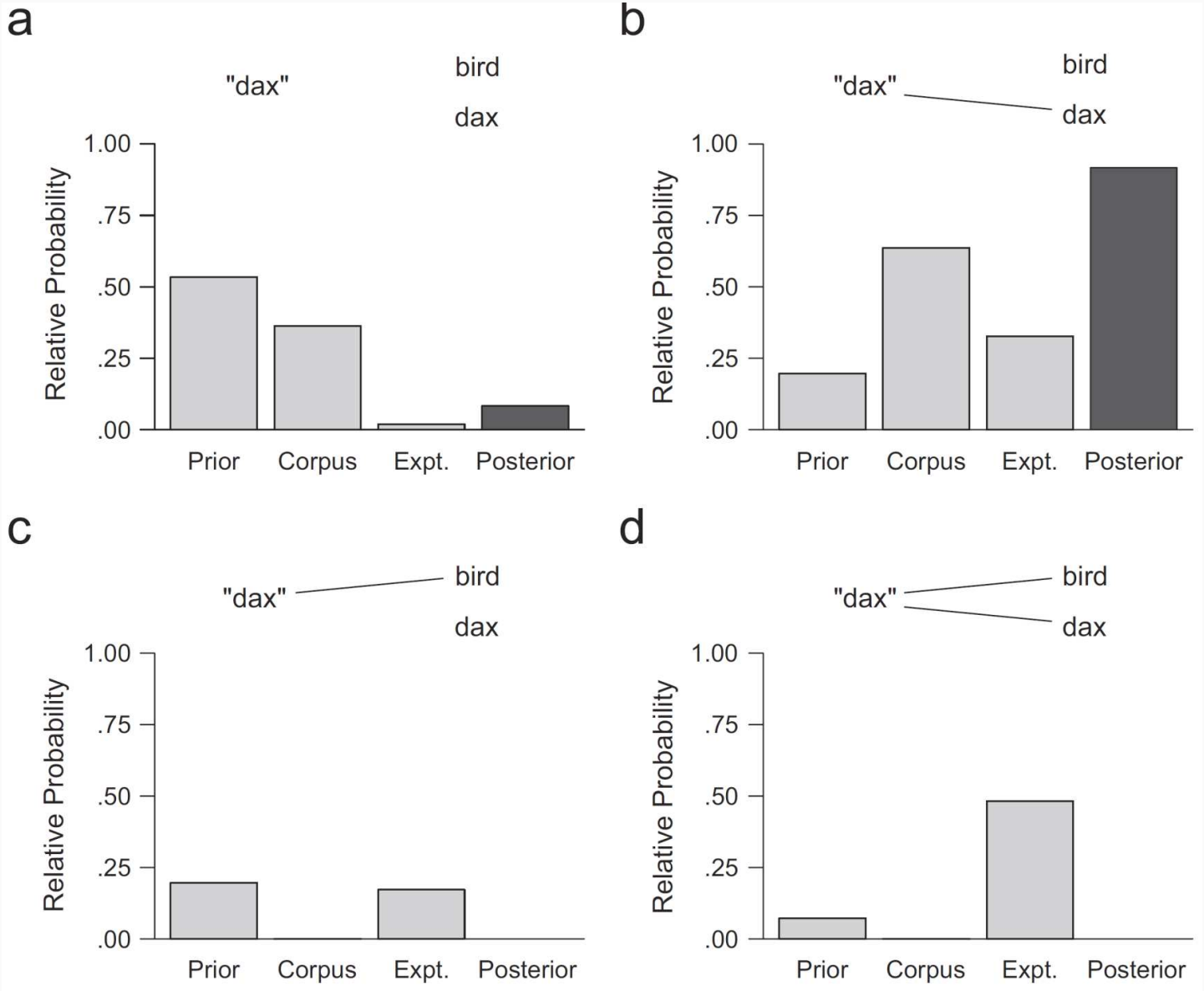
Results: Mutual Exclusivity

Children as young as 16 months map novel words to novel objects:



- some researchers have postulated a principle of mutual exclusivity to account for this;
- but it could also be general pragmatic principles at work;
- is mutual exclusivity learned or innate?

Results: Mutual Exclusivity



Frank et al. (2009)

Results: Mutual Exclusivity

The model is able to capture mutual exclusivity:

- mapping "dax" to BIRD is unlikely:
  - highly coincidental that no other BIRDs are "dax";
  - corpus likelihood is low
- prior favours not mapping "dax" to anything, but this lowers the probability of the data (corpus and experiment).
- Many of the other models also predict mutual exclusivity, suggesting no special principle is needed.

This example also shows that the model captures *one-trial learning.*

Discussion

Strengths:

- model combines cross-situational and social/intentional learning ("joint learning")
- more accurate learning of lexicon and referents than previous models;
- explains experimental phenomena without special principles.

Weaknesses:

Discussion

Strengths:

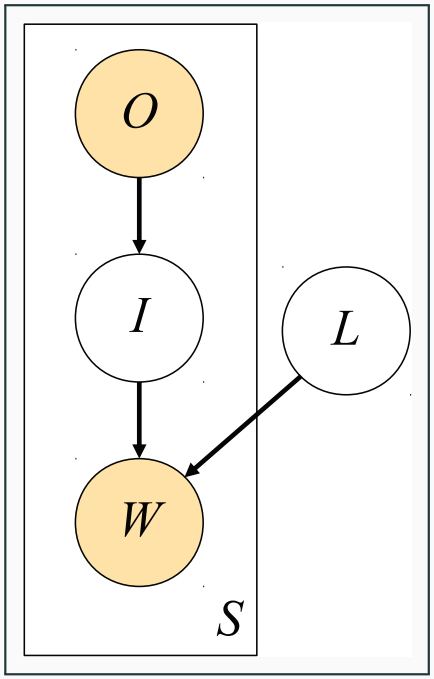
- model combines cross-situational and social/intentional learning ("joint learning")
- more accurate learning of lexicon and referents than previous models;
- explains experimental phenomena without special principles.

Weaknesses:

- only tested on very small corpus;
- only deals with concrete nouns;
- no model of syntax.
- intention learning is not grounded to observed social cues.

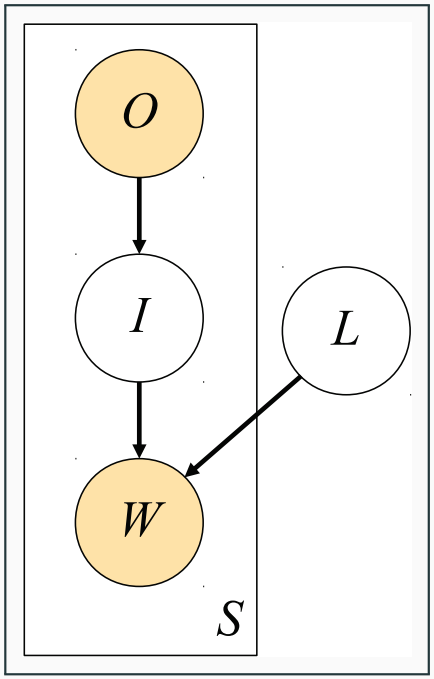
Intentions and Social Cues

Intentions



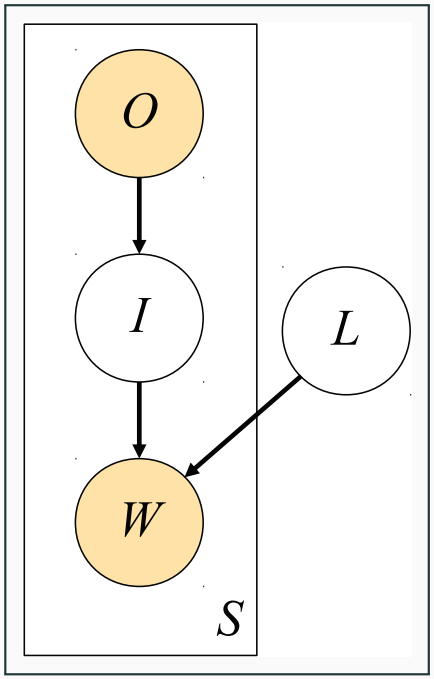
What is ‘Intention’ latent variable really doing?

Intentions



Arguably: Mostly making non-referential word use (“NULL alignment”) more likely.

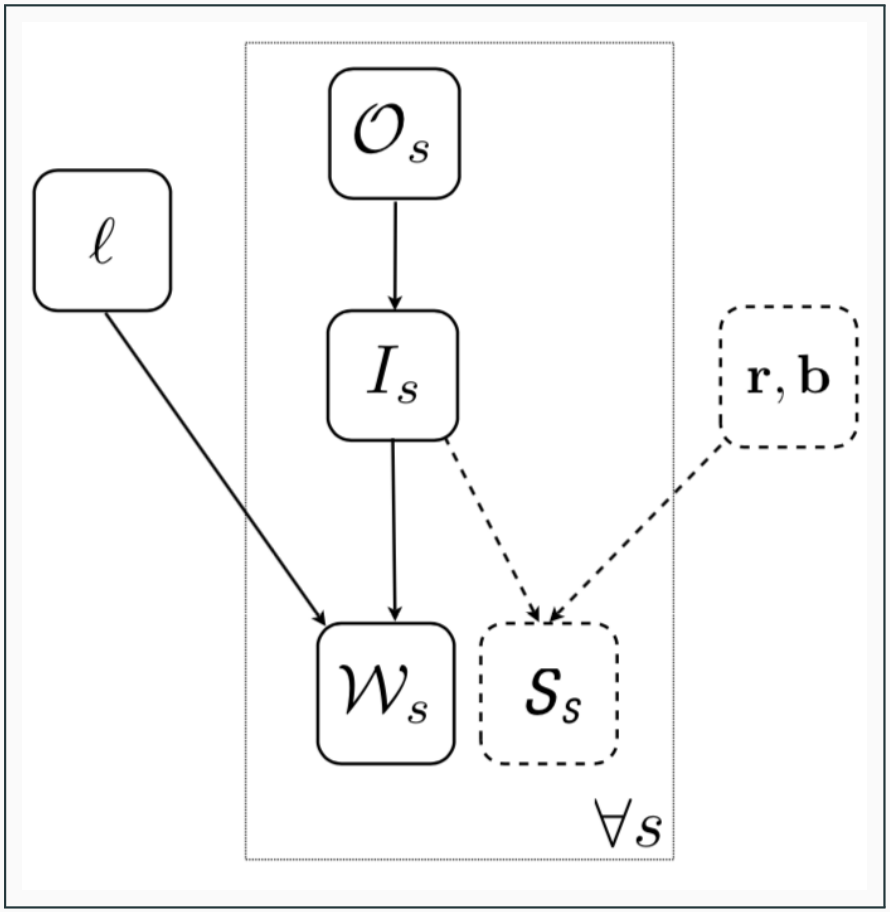
Intentions



How could we add actual social cues to the model?

e.g. child's, parent's eye gaze, hands

Frank et al. (2008)
A Bayesian Framework for Cross-Situational Word-Learning



Frank et al. (2008)

Hyperparameters: r is 'relevance'; b is 'base rate'.

Social cues are also generated from intentions: $P(S|I, r, b)$

New likelihood:

$$P(W, S|L) \propto \prod_{s \in D} \sum_{I_s \subseteq O_s} P(I_s|O_s) P(W_s|I_s, L) P(S_s|I)$$

Summary

- Infants learn words using multiple cues and strategies:
  - Statistical learning, by tracking word and objects over time using cross-situational learning
  - Social cues, such as gaze and pointing, to disambiguate within a single context
- Frank et al. model includes these joint strategies
- Infers better lexicon than pure alignment/co-occurrence model
- Captures mutual exclusivity behaviour

Summary

- Infants learn words using multiple cues and strategies:
  - Statistical learning, by tracking word and objects over time using cross-situational learning
  - Social cues, such as gaze and pointing, to disambiguate within a single context
- Frank et al. model includes these joint strategies
- Infers better lexicon than pure alignment/co-occurrence model
- Captures mutual exclusivity behaviour

Questions?

Next time

- For Thursday: read Perfors' 2011 tutorial.
  - Minimally: Focus on Sections 1+2 (skip 2.1) and Section 3
  - Read sections 4 to end later (or now!)
- Thursday: Hierarchical Bayesian Models
- Paper: *Kemp, Perfors and Tenenbaum (2007). Learning overhypotheses with hierarchical Bayesian models*
- How to *infer* more informed priors, in order to build models that can capture more complexity
- e.g. to learn that some objects are more likely to be intended than others (instead of having a flat prior)
- or to learn good (distributions over) hyperparameter values (e.g. distribution over α for lexicon prior)