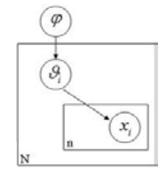
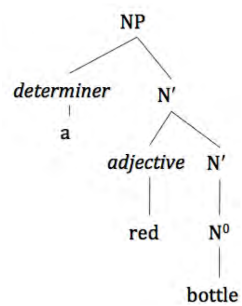
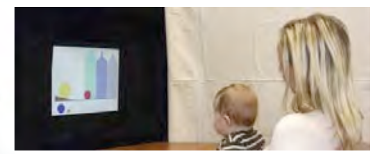
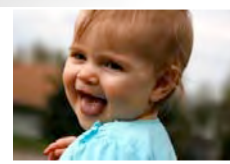
Knowing where to look:
Identifying what children need to make syntactic generalizations









Lisa Pearl
University of California, Irvine

Computation of Language Laboratory
UC Irvine

Oct 10, 2013: Cognition and Language Workshop

Stanford University

The process of language learning

The process of language learning

Given the available input…

Look at that kitty!
There's another one.

Input

Where did he hide?
What happened?

The process of language learning

Given the available input, information processing done by human minds…

Look at that kitty!
There's another one.

Input

Where did he hide?
What happened?

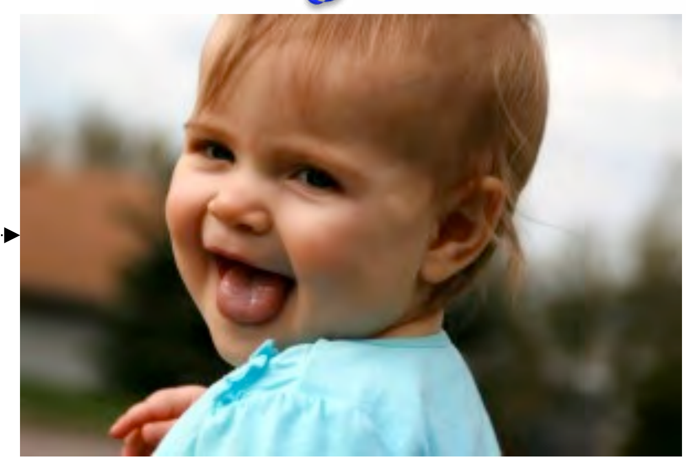
The process of language learning

Given the available input, information processing done by human minds to build a system of linguistic knowledge…

Look at that kitty!
There's another one.

Input

Where did he hide?
What happened?



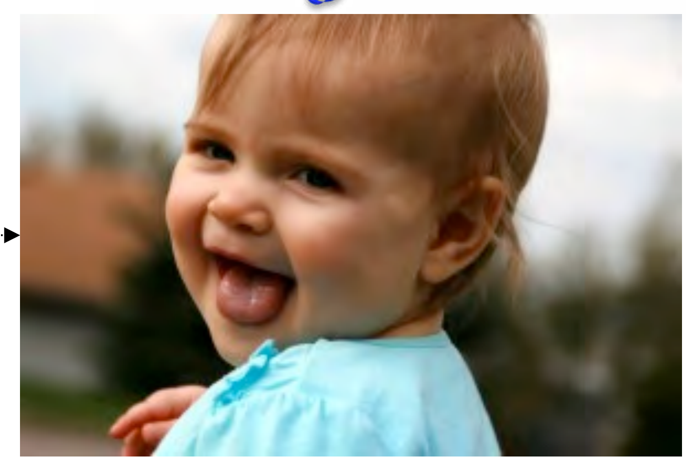
The process of language learning

Given the available input, information processing done by human minds to build a system of linguistic knowledge whose output we observe

Look at that kitty!
There's another one.

Input

Where did he hide?
What happened?



Output

Where's the kitty?
That one's really cute.

Making generalizations

Why can learning be tricky?

One issue: **Induction problems**

There are often many ways to generalize beyond the input, and most of them aren't right.

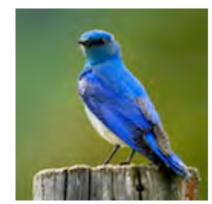
Making generalizations

Why can learning be tricky?

One issue: **Induction problems**

There are often many ways to generalize beyond the input, and most of them aren't right.

"birdie" =



data encountered

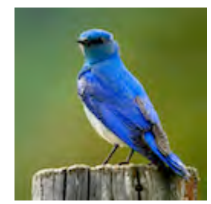
Making generalizations

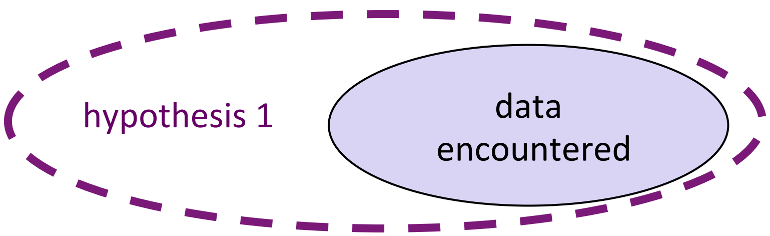
Why can learning be tricky?

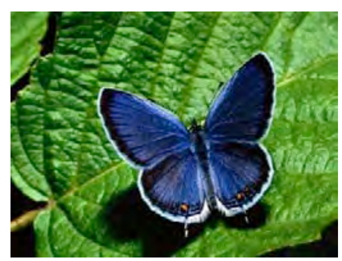
One issue: **Induction problems**

There are often many ways to generalize beyond the input, and most of them aren't right.

"birdie" =







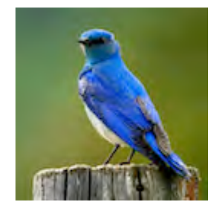
Making generalizations

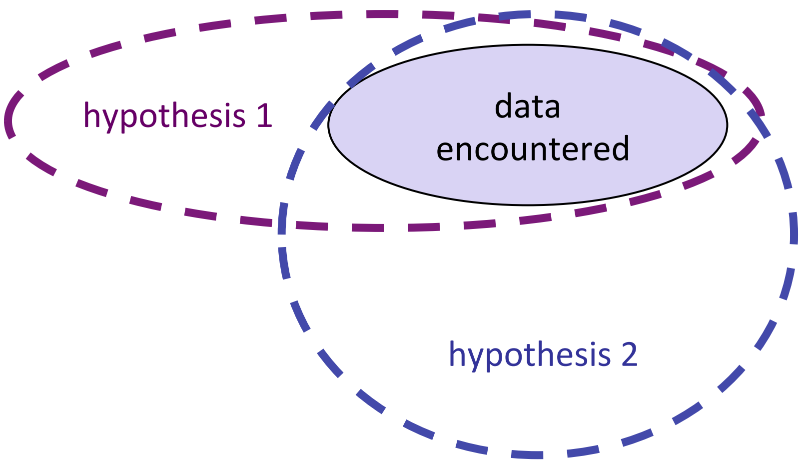
Why can learning be tricky?

One issue: **Induction problems**

There are often many ways to generalize beyond the input, and most of them aren't right.

"birdie" =

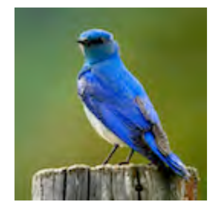


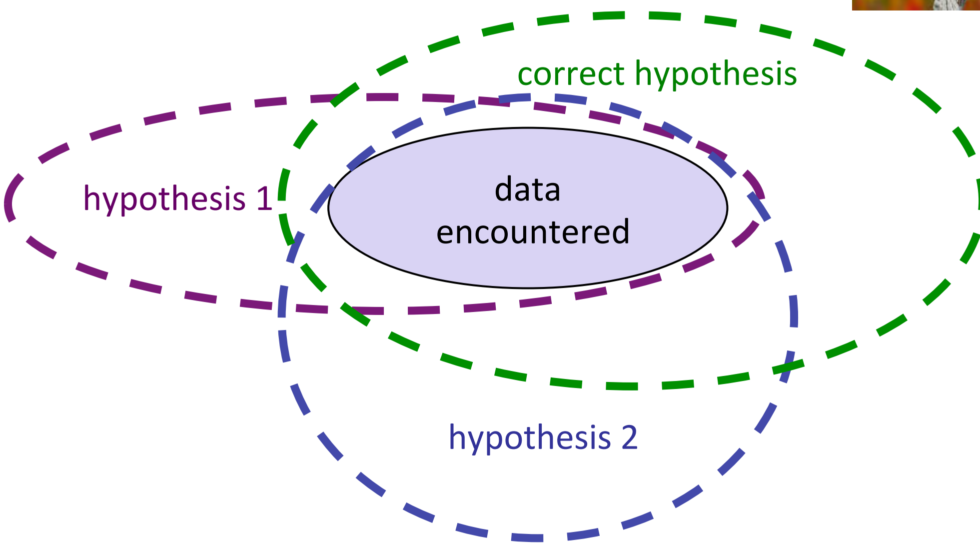


Making generalizations

Why can learning be tricky?
One issue: **Induction problems**
There are often many ways to generalize beyond the input, and most of them aren't right.

"birdie" =



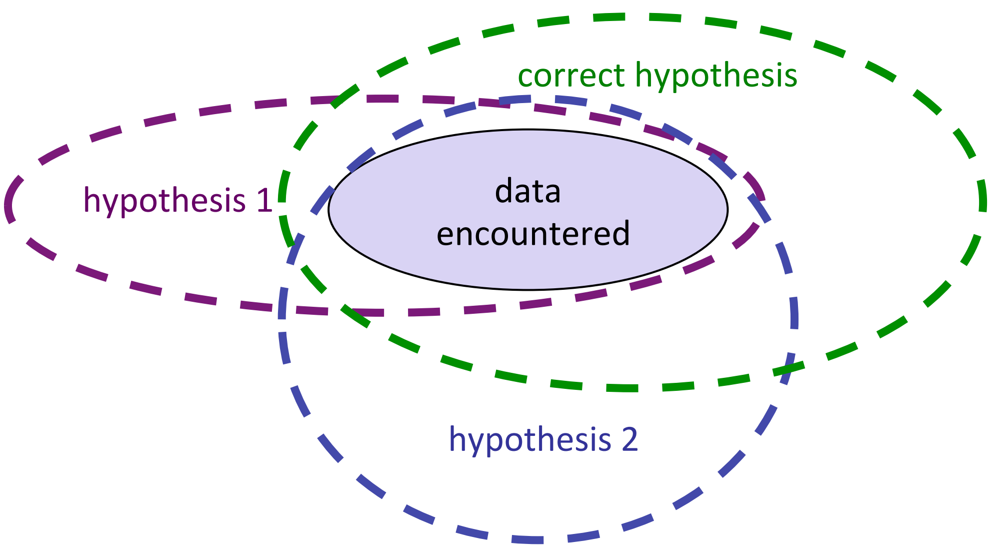


Making generalizations

Why can learning be tricky?

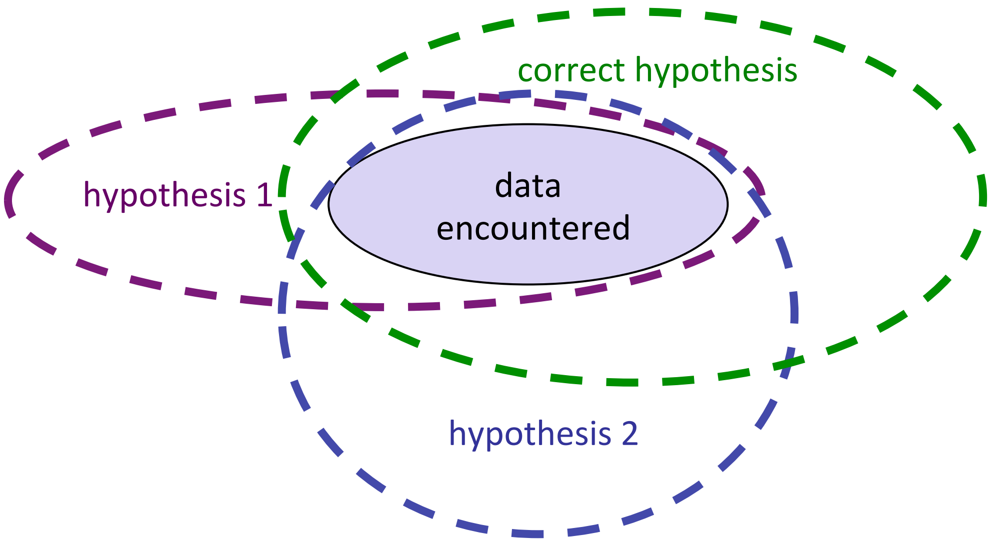
One issue: **Induction problems**

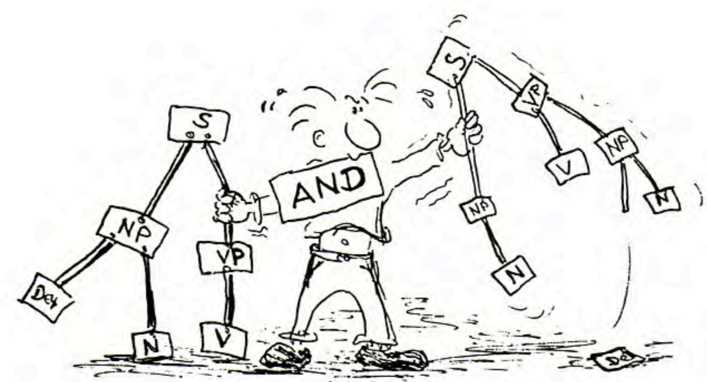
This has sometimes been called the Poverty of the Stimulus, the Logical Problem of Language Acquisition, or Plato's Problem.



Making generalizations

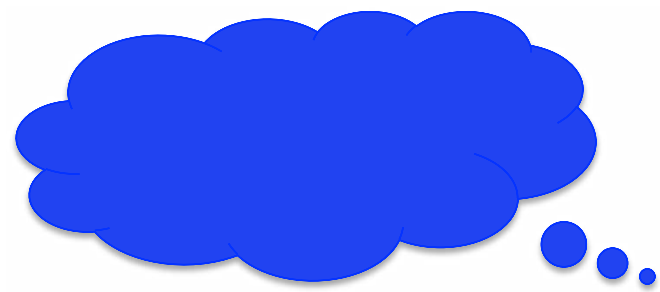
Though induction problems occur for all kinds of knowledge acquisition, today's focus = syntactic knowledge.

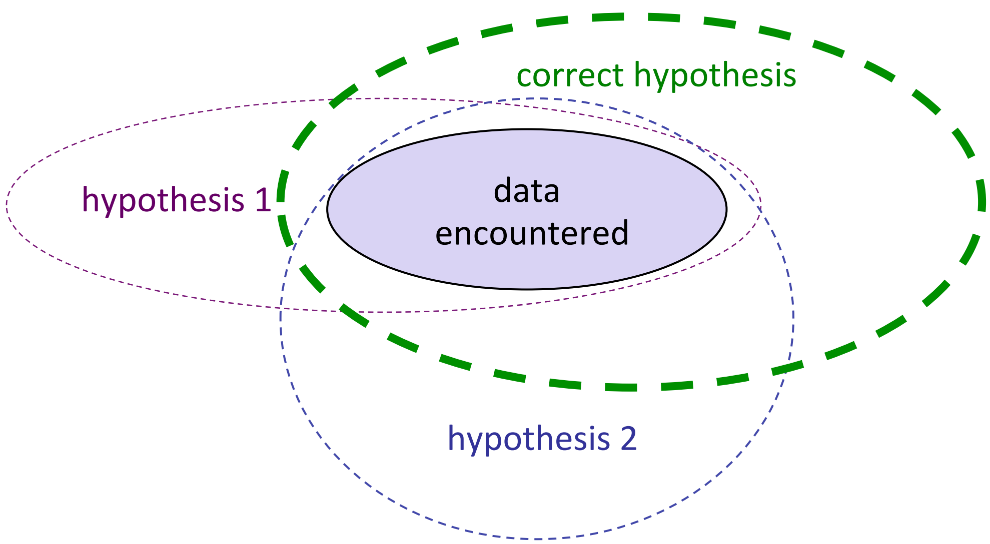




Making generalizations

One solution to induction problems:
Helpful learning strategies that guide the types of generalizations learners make.





Making syntactic generalizations

Previous suggestions for how children make specific syntactic generalizations tend to involve learning strategies containing very specific (and often linguistic) prior knowledge.

Making syntactic generalizations

Previous suggestions for how children make specific syntactic generalizations tend to involve learning strategies containing very specific (and often linguistic) prior knowledge.

Some examples:

- *Syntactic islands*: Knowing that certain linguistic dependencies are limited to crossing no more than a single specific, abstract linguistic structure

  (Chomsky 1973, Huang 1982, Lasnik & Saito 1984)

Wh … $[_{BN2}$ … $[_{BN1}$ … ___]]

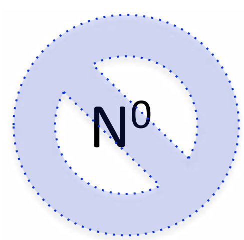
Making syntactic generalizations

Previous suggestions for how children make specific syntactic generalizations tend to involve learning strategies containing very specific (and often linguistic) prior knowledge.

Some examples:

- *English anaphoric one*: Knowing certain grammatical category assignments are illicit for particular kinds of words in the language

  (Baker 1978)

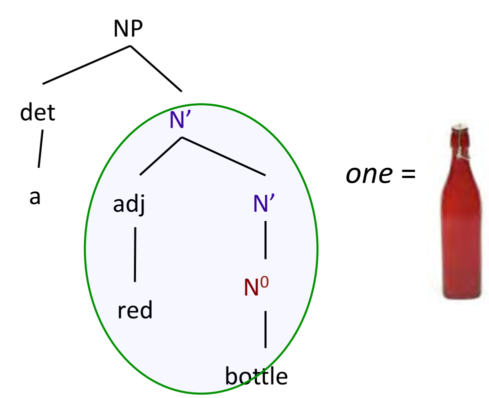


Making syntactic generalizations

Recent investigations:
Demonstrating for these two case studies that learning strategies involving less specific knowledge are sufficient.

- Syntactic islands
  (Pearl & Sprouse 2013a, Pearl & Sprouse 2013b)

- English anaphoric *one*
  (Pearl & Mis 2011, Pearl & Mis 2013, Pearl & Mis under review)



Making syntactic generalizations

Recent investigations:
Demonstrating for these two case studies that learning strategies involving less specific knowledge are sufficient.

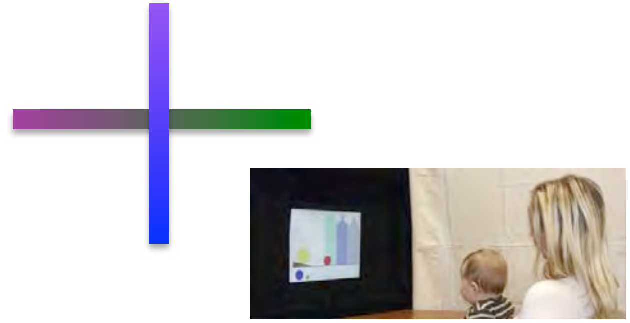
Recurring themes:

(1) Broadening the set of data perceived as informative with indirect positive evidence

(2) Matching the empirical data we have about the target knowledge state via observable behavior

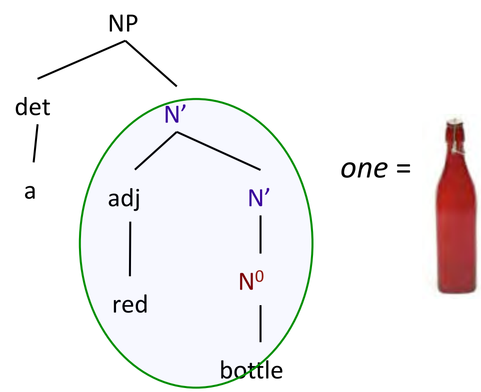
Today's plan

I. Recurring themes: evidence types & target states



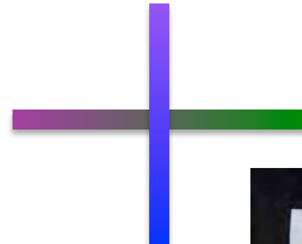
II. Defining the learning task so we can figure out what's needed to solve it

III. Case study: English anaphoric *one*



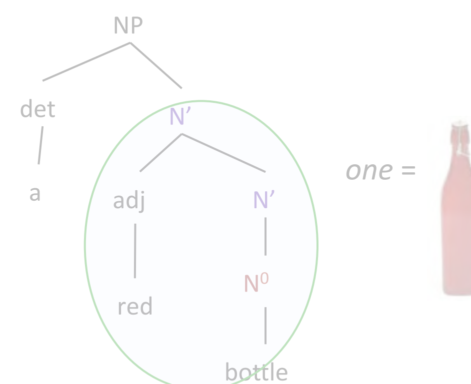
Today’s plan

I. Recurring themes: evidence types & target states



II. Defining the learning task so we can figure out what’s needed to solve it

III. Case study: English anaphoric *one*



Types of evidence for making generalizations

Some relevant distinctions:

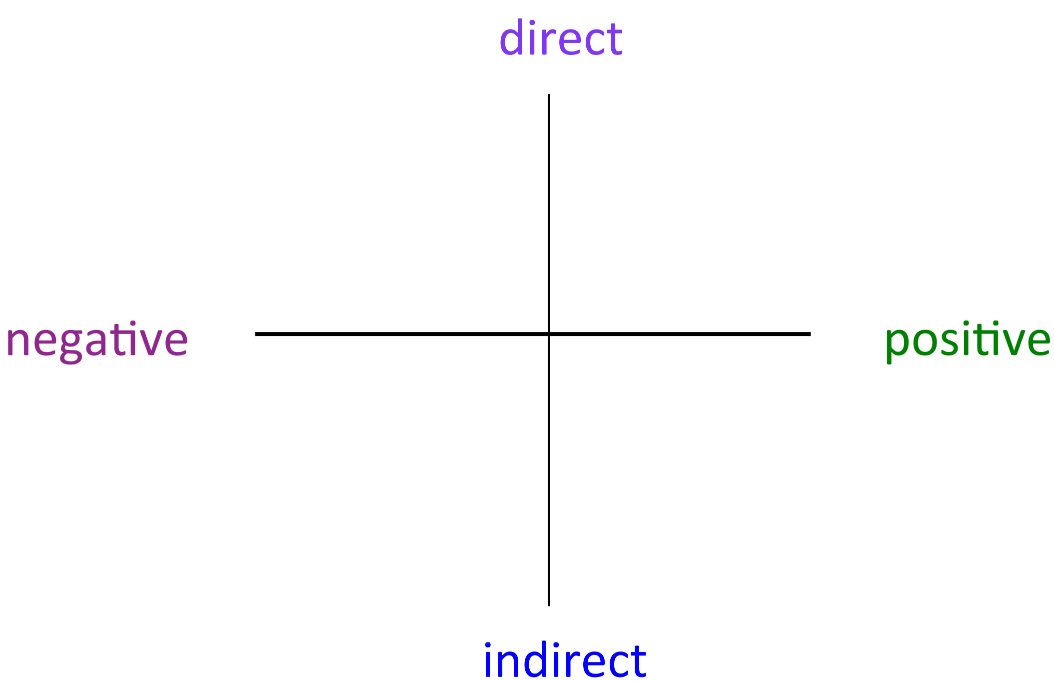
(i) positive vs. negative: Is the evidence about items that are present or items that are absent from the language?

negative ———————————— positive

Types of evidence for making generalizations

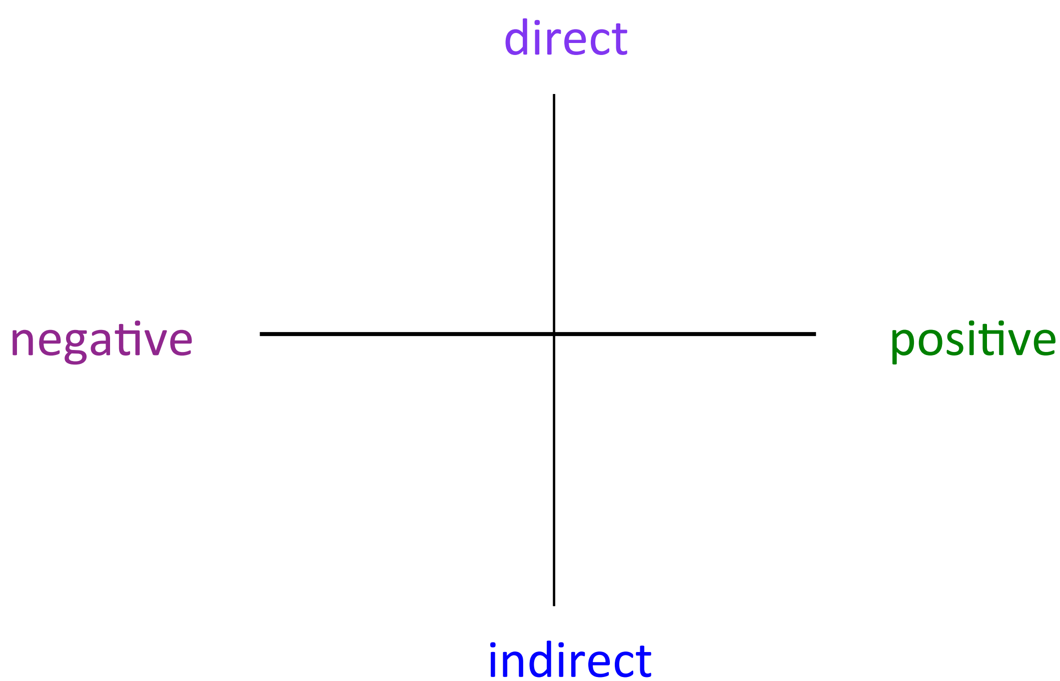
Some relevant distinctions:
(i) positive vs. negative: Is the evidence about items that are present or items that are absent from the language?

(ii) direct vs. indirect: Is it certain that the items are (un)grammatical, or does it require inference on the part of the learner?



Types of evidence for making generalizations

Evidence types:



Utterances:

(i) *Jack has a red bottle but he wants another one.*

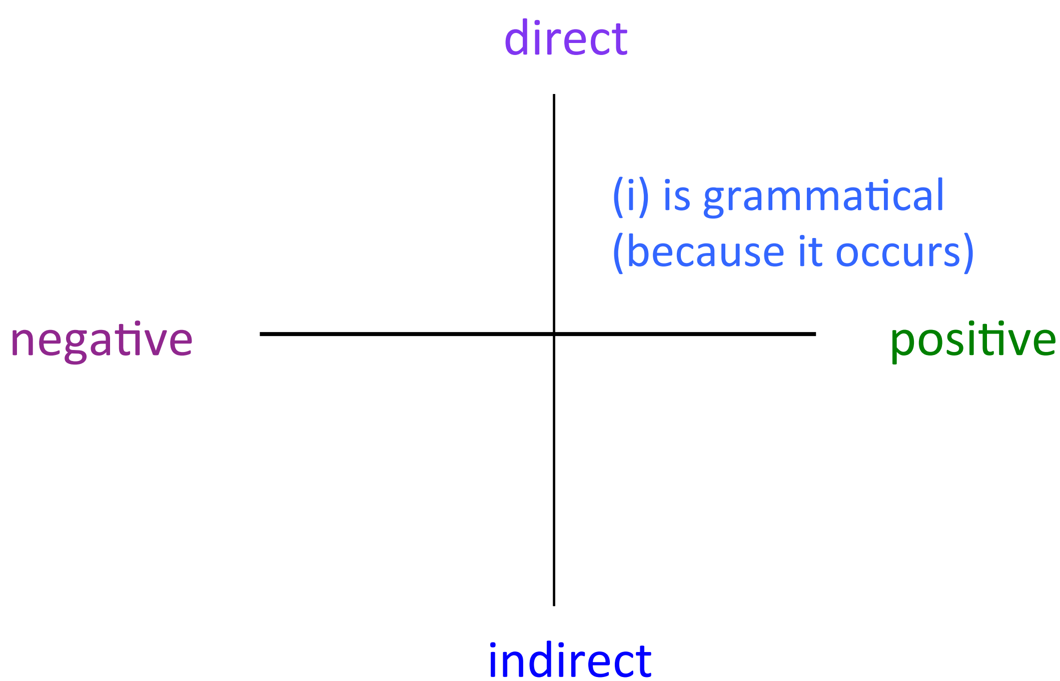
(ii) **Jack sat by the side of the building and Lily sat by the one of the road.*

(iii) *Jack has a red bottle and Lily wants it.*

Types of evidence for making generalizations

Evidence types:

direct positive evidence (traditionally assumed to be available)



Utterances:

(i) *Jack has a red bottle but he wants another one.*

(ii) **Jack sat by the side of the building and Lily sat by the one of the road.*

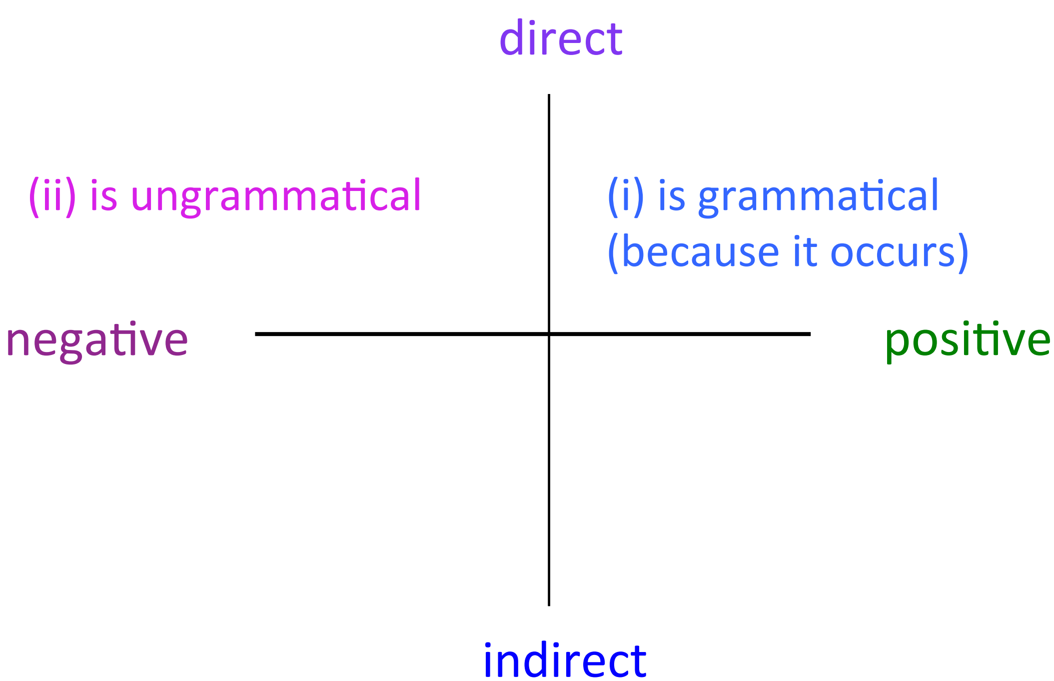
(iii) *Jack has a red bottle and Lily wants it.*

Types of evidence for making generalizations

Evidence types:

direct positive evidence (traditionally assumed to be available)

direct negative evidence (typically assumed to be unavailable or ignored)



Utterances:

(i) Jack has a red bottle but he wants another one.

*(ii) *Jack sat by the side of the building and Lily sat by the one of the road.*

(iii) Jack has a red bottle and Lily wants it.

Types of evidence for making generalizations

Evidence types:

direct positive evidence (traditionally assumed to be available)

direct negative evidence (typically assumed to be unavailable or ignored)

indirect negative evidence (assumed to potentially be available, usually for a statistical learner)

direct

(ii) is ungrammatical

(i) is grammatical
(because it occurs)

negative

positive

(ii) has not occurred yet, so
maybe it's ungrammatical

indirect

Utterances:

(i) *Jack has a red bottle but he wants another one.*

(ii) **Jack sat by the side of the building and Lily sat by the one of the road.*

(iii) *Jack has a red bottle and Lily wants it.*

Types of evidence for making generalizations

Evidence types:

direct positive evidence (traditionally assumed to be available)

direct negative evidence (typically assumed to be unavailable or ignored)

indirect negative evidence (assumed to potentially be available, usually for a statistical learner)

indirect positive evidence (not often explicitly recognized for syntactic induction problems, but potentially available)

direct

(ii) is ungrammatical

(i) is grammatical (because it occurs)

negative

positive

(ii) has not occurred yet, so maybe it's ungrammatical

(iii) occurs, so maybe (i) is grammatical and maybe (ii) is ungrammatical

indirect

Utterances:

(i) *Jack has a red bottle but he wants another one.*

(ii) **Jack sat by the side of the building and Lily sat by the one of the road.*

(iii) *Jack has a red bottle and Lily wants it.*

Indirect positive evidence

Indirect positive evidence is related to the ideas behind linguistic parameters and Bayesian overhypotheses. Both allow data besides those about the specific items of interest to be deemed informative.

Indirect positive evidence

Indirect positive evidence is related to the ideas behind linguistic parameters and Bayesian overhypotheses. Both allow data besides those about the specific items of interest to be deemed informative.

Linguistic parameters:
Data about knowledge$_1$ can help set the linguistic parameter, which in turn helps determine knowledge$_2$.

Linguistic parameter

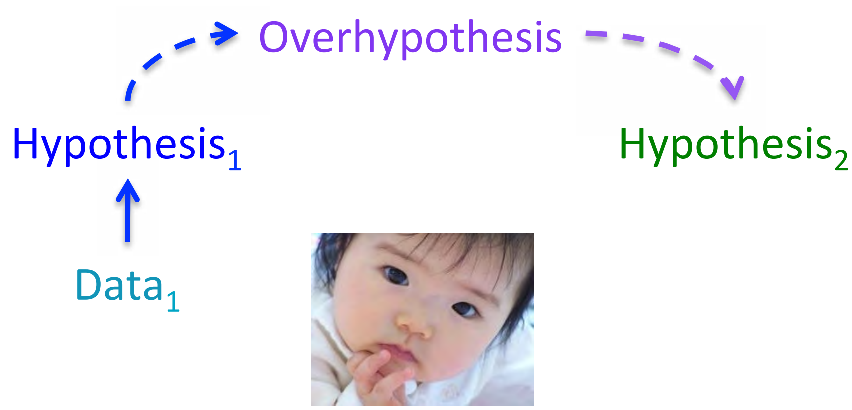
Knowledge$_1$ Knowledge$_2$

Data$_1$

Indirect positive evidence

Indirect positive evidence is related to the ideas behind linguistic parameters and Bayesian overhypotheses. Both allow data besides those about the specific items of interest to be deemed informative.

Overhypotheses:
Data about hypothesis_1 can help specify the overhypothesis, which in turn helps determine hypothesis_2.

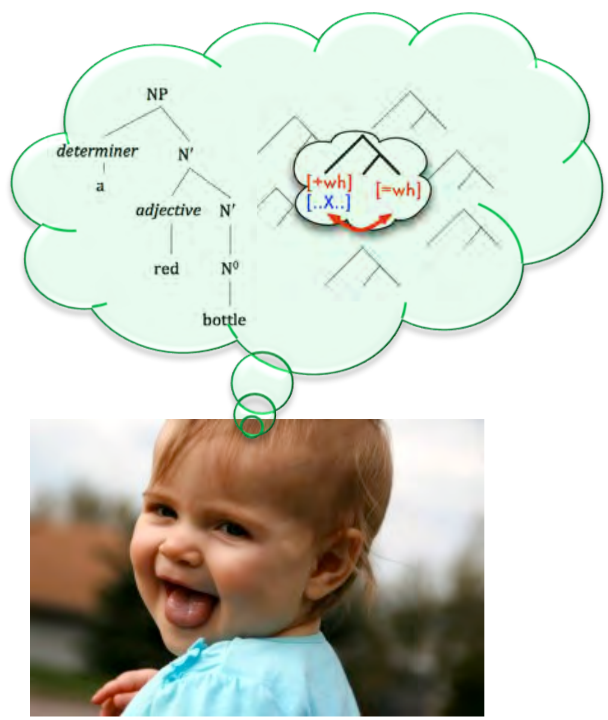


Empirically defining the target state

The goal of learning is usually described as reaching a certain target knowledge state.

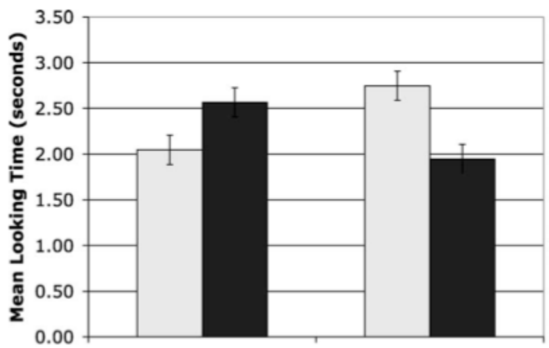
Ex: making the correct syntactic generalization from the data

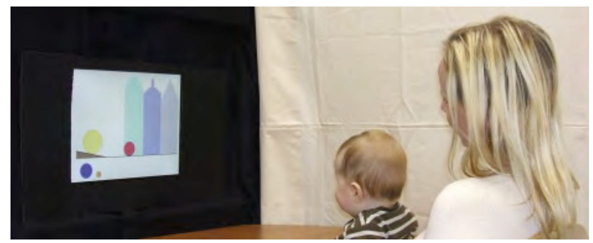
Problem: Knowledge states aren't easily observable.



Empirically defining the target state

Solution: Experiments allow us to observe behavior generated by an individual's knowledge state.





Can we deduce the underlying knowledge state that generated this target behavior?

If so, the target behavior is a good proxy for the target knowledge state.

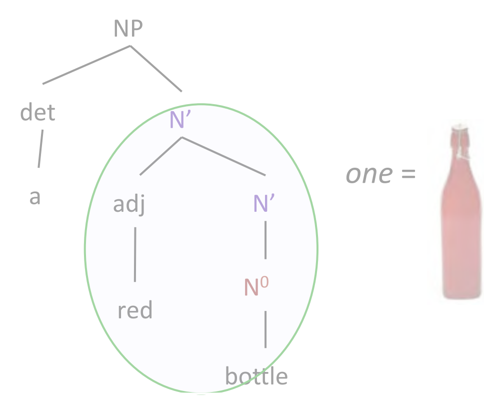
Updated goal: Determine how learners can make the syntactic generalizations that lead to the observed target behavior.

Today's plan

I. Recurring themes: evidence types & target states

II. Defining the learning task so we can figure out what's needed to solve it

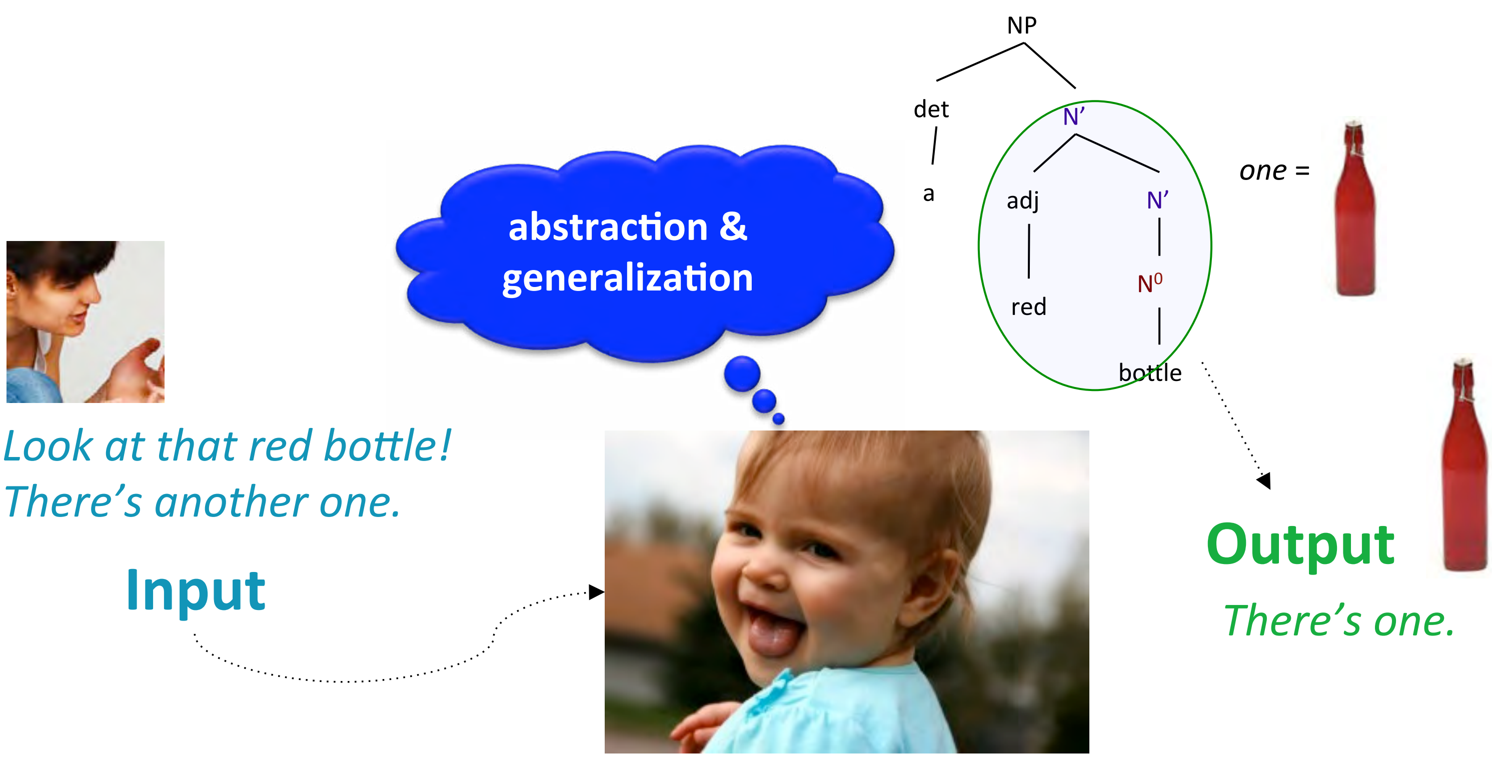
III. Case study: English anaphoric *one*



Components of the learning task

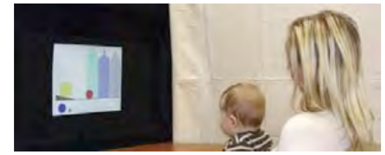
The language learning process has some well-defined pieces already: input, abstraction/generalization, inferred knowledge, and observable output.

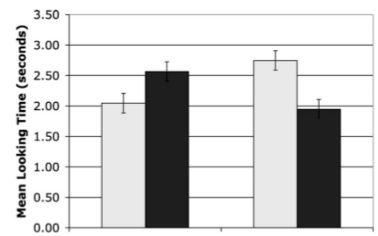
These correspond to major components of the learning task.

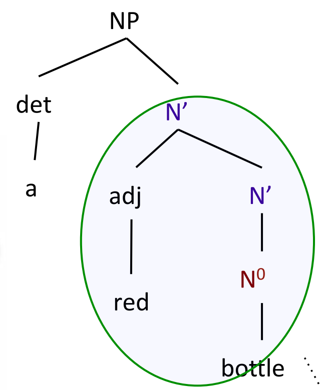


Components of the learning task

Target state: The knowledge children are trying to attain, which we can gauge through their observable behavior.





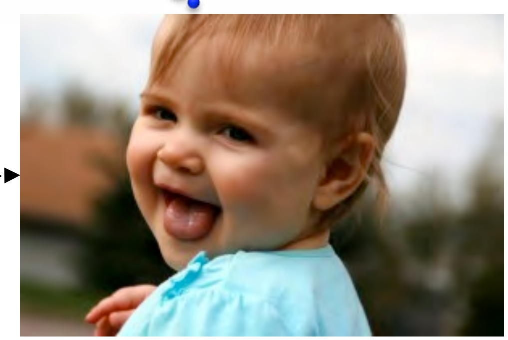


one =

Target state

Look at that red bottle!
There's another one.

Input



Output

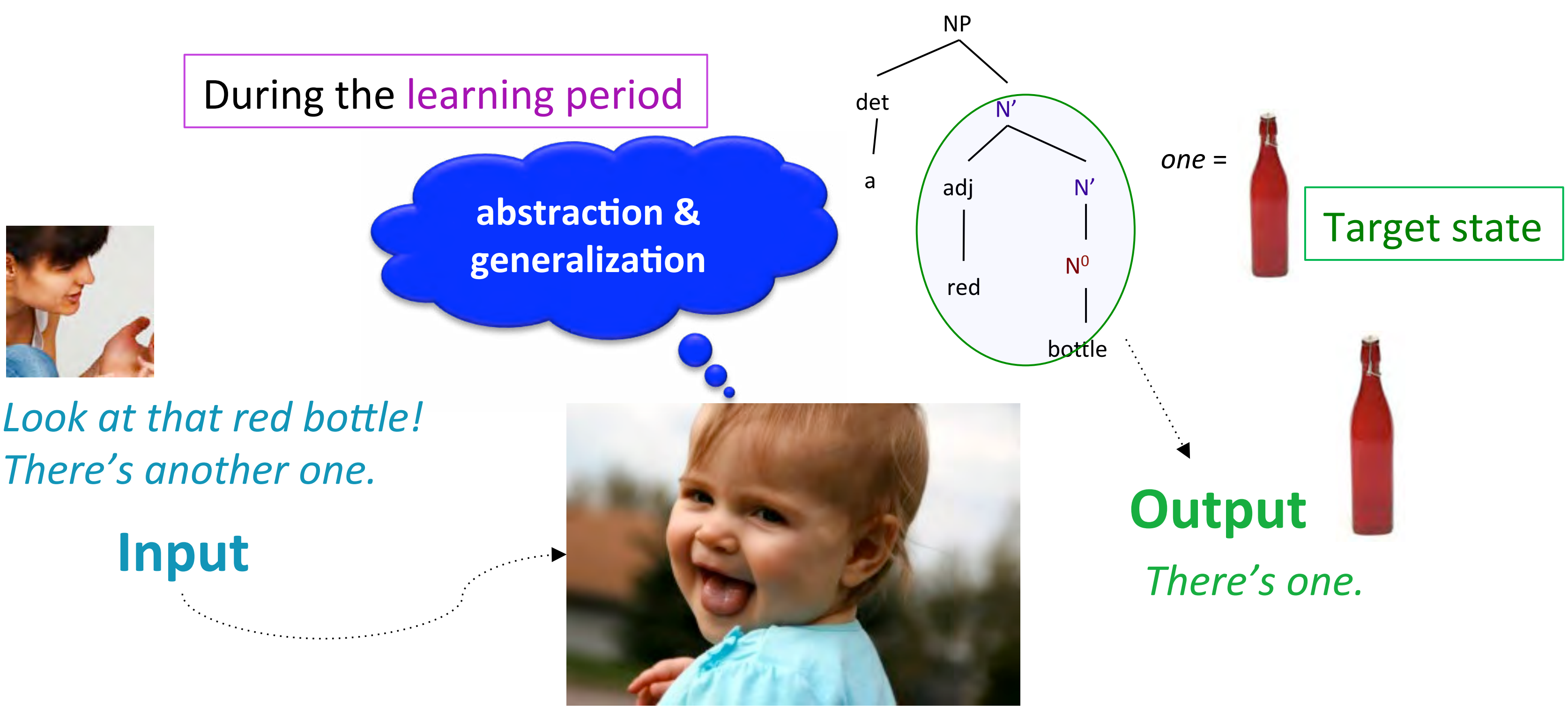
There's one.

Components of the learning task

When abstraction & generalization happen

Learning period: How long children have to reach the target state.

- Can be defined by time (ex: 4 months) or quantity of data encountered (ex: 36,500).

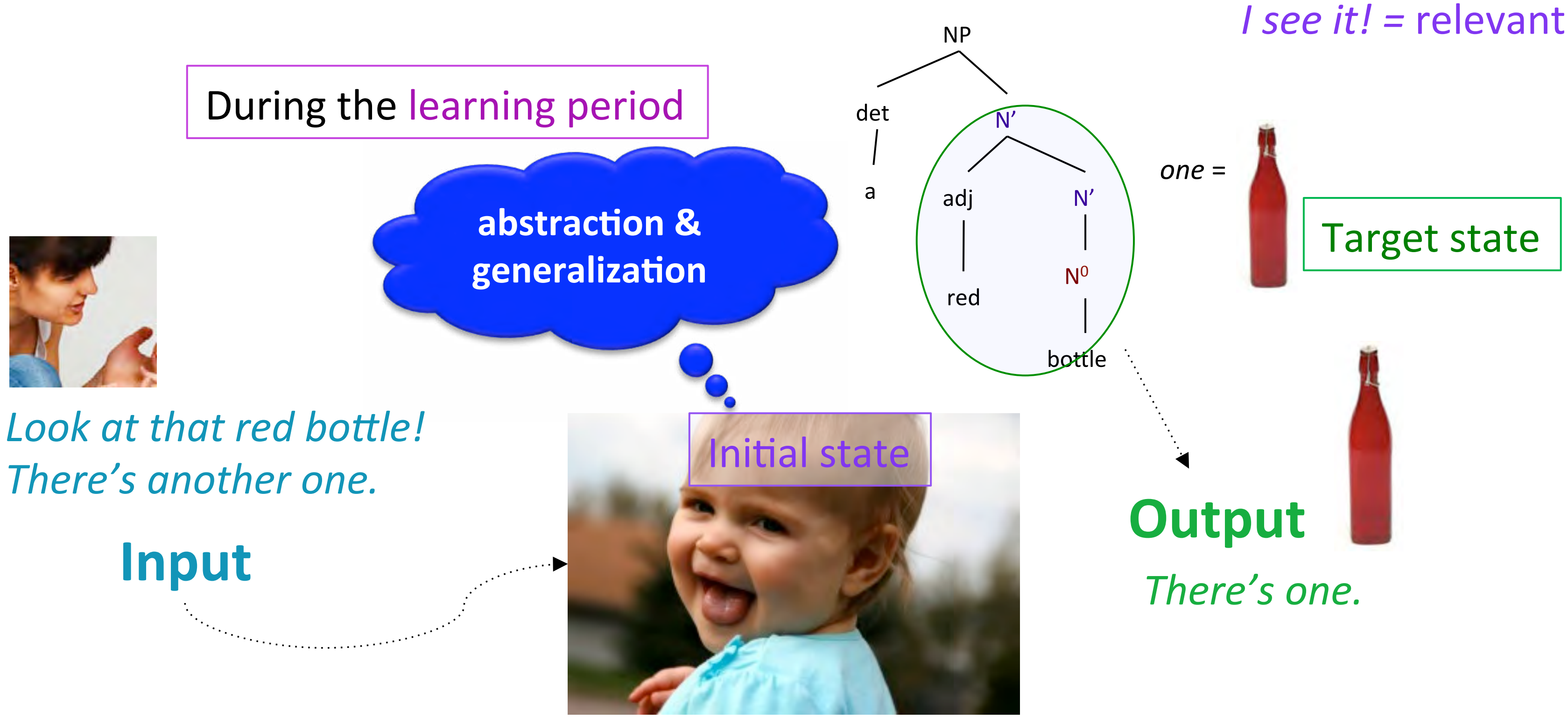


Components of the learning task

How abstraction & generalization happen

Initial state: The knowledge, capabilities, and biases children have.

- prior knowledge for constraining generalizations (ex: knowing grammatical categories) N^0, N', NP, …
- prior learning capabilities (ex: tracking frequency information) N' = N' + 1
- learning biases (ex: being sensitive to certain information in the input) *I see it!* = relevant



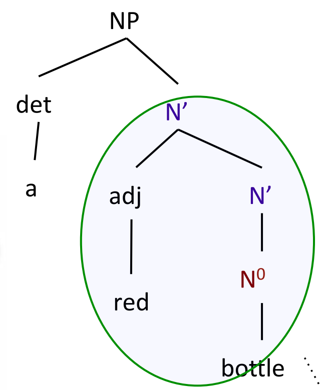
Components of the learning task

Data intake: Data perceived as relevant for learning (Fodor 1998).

Often a subset of the available input, winnowed down by the learner's biases.

During the learning period

abstraction & generalization



one =

Target state

Look at that red bottle!
There's another one.

Input

Data intake

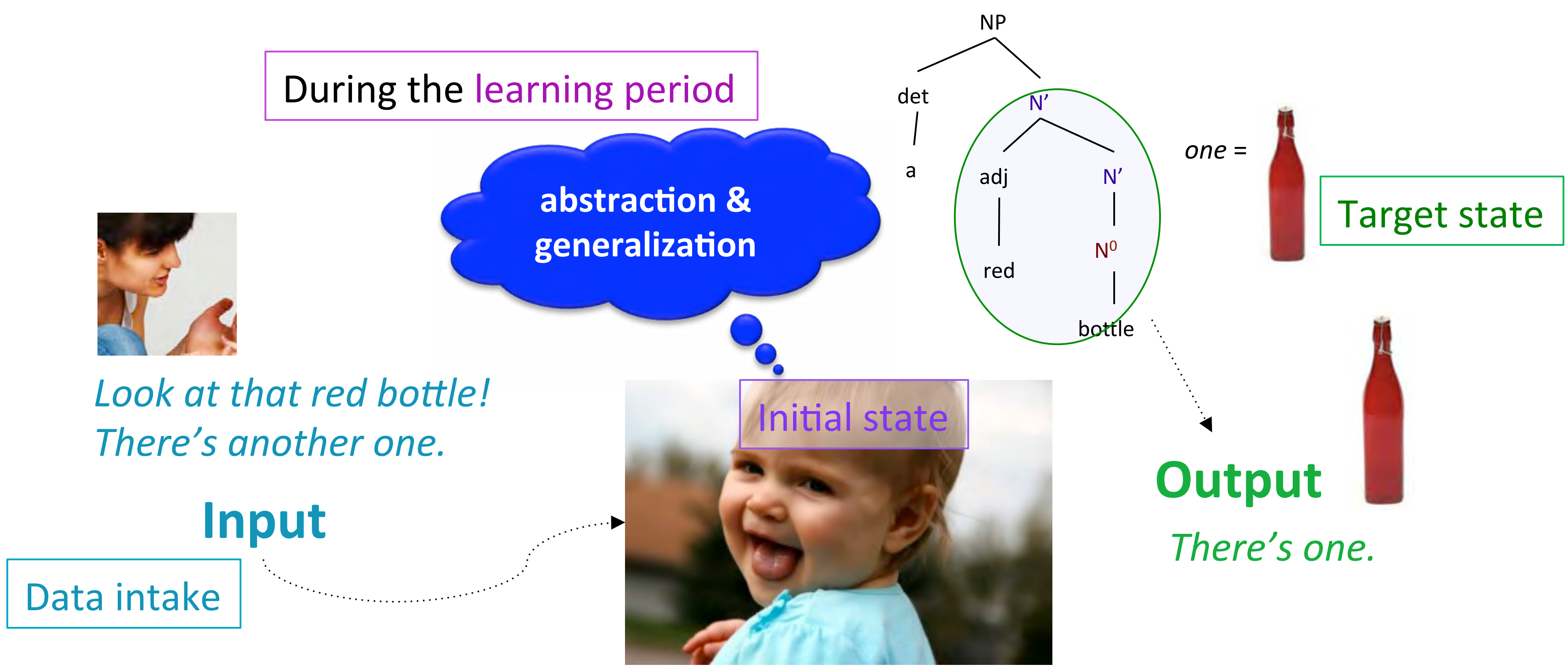
Output

There's one.

Components of the learning task

Learning task definition:

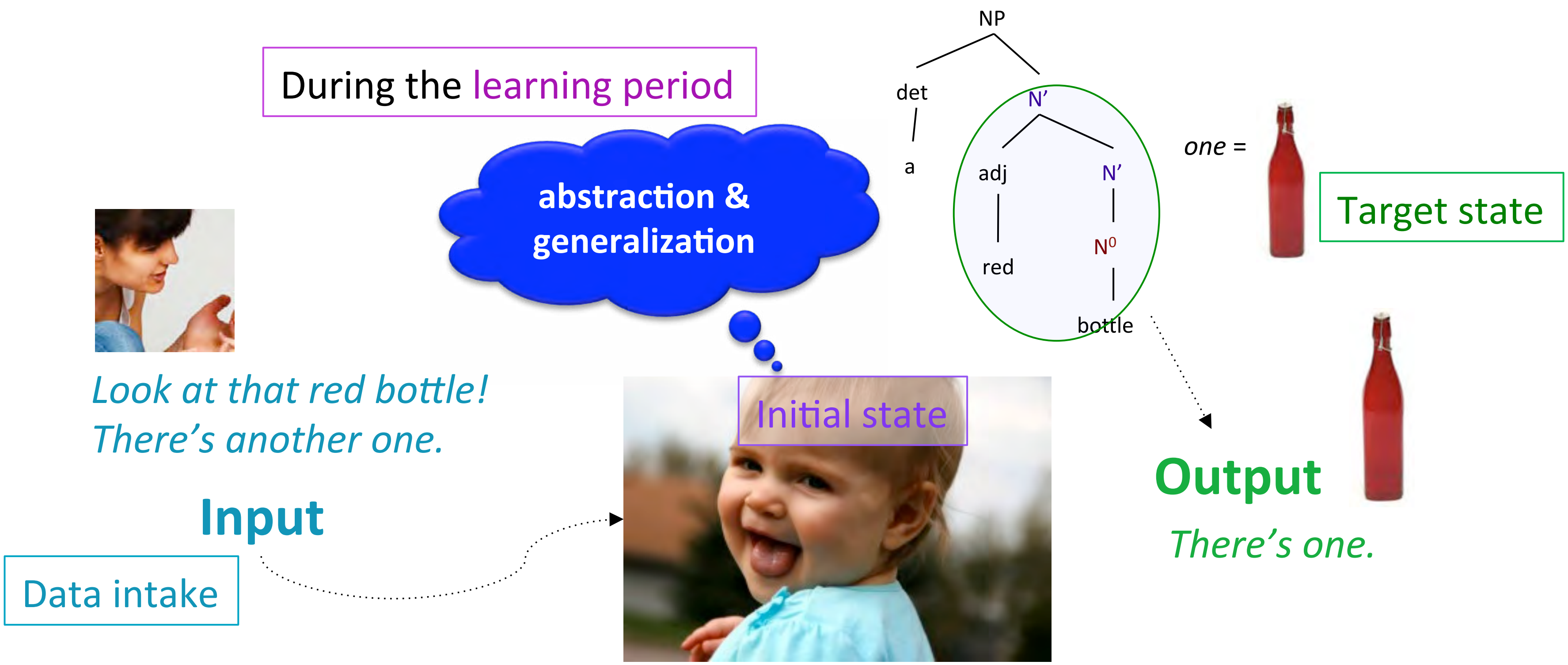
Given a specific initial state, a learner must use the data intake to reach the target state by the end of the learning period.



Components of the learning task

We can then use this definition to explore potential learning strategies.

Goal: Learn the appropriate what by the appropriate when using some kind of cognitively plausible how and the available input.



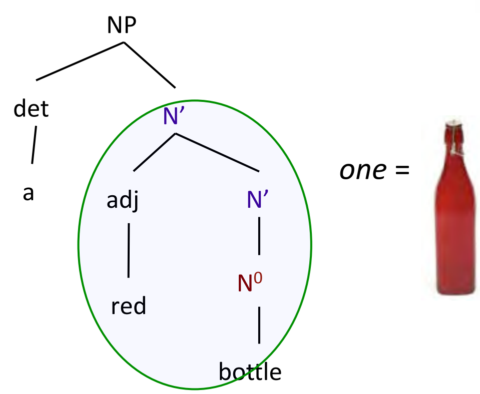
To identify effective learning strategies for making syntactic generalizations, we need to draw on a variety of research methods to specify the components of the learning task.

To identify effective learning strategies for making syntactic generalizations, we need to draw on a variety of research methods to specify the components of the learning task.

Theoretical methods:

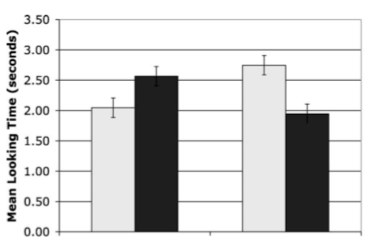
What the knowledge is

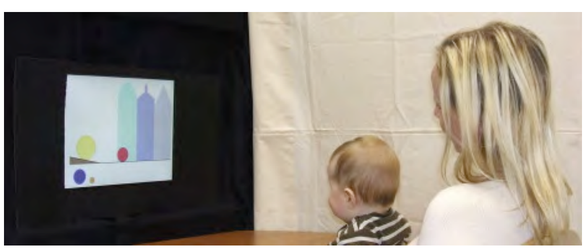
[target (knowledge) state]



To identify effective learning strategies for making syntactic generalizations, we need to draw on a variety of research methods to specify the components of the learning task.

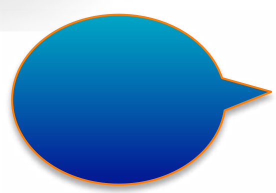
Experimental methods:
When knowledge is acquired (as evidenced by behavior), what the **input** looks like, & plausible capabilities underlying **how** acquisition works
[learning period, target (behavior) state, data intake, initial state]





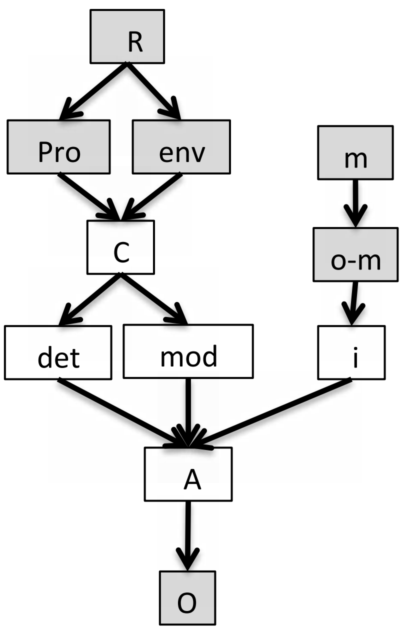
$p(\mathrm{H1}|$  $)$

$\propto p($  $|\,\mathrm{H1})\ p(\mathrm{H1})$



To identify effective learning strategies for making syntactic generalizations, we need to draw on a variety of research methods to specify the components of the learning task.

Computational methods:
Biases and capabilities that are useful for **how** children acquire knowledge + quantitative analysis of input
[initial state, data intake]



Learning strategies

When we find a successful learning strategy, this is an existence proof that the syntactic generalization is possible using the learning biases, knowledge, and capabilities comprising that strategy.

This identifies useful learning strategy components, which we can then examine to see where they might come from.

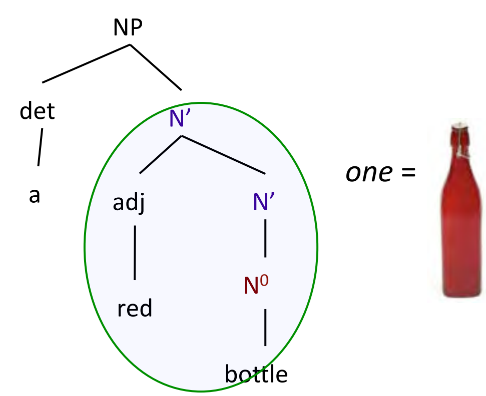
Knowledge 1
Knowledge 2
Capability 1
Bias 1
Bias 2
Bias 3
…

Today's plan

I. Recurring themes: evidence types & target states

II. Defining the learning task so we can figure out what's needed to solve it

III. Case study: English anaphoric *one*



English anaphoric *one*

Look - a red bottle!

Pearl & Mis submitted

English anaphoric *one*

Look - a red bottle!

Look – another *one*!

Pearl & Mis submitted

English anaphoric *one*

Look - a red bottle!

red bottle

Look – another *one*!

Process: First determine the antecedent of *one* (what expression *one* is referring to).
→"red bottle"

Pearl & Mis submitted

English anaphoric *one*

Look - a red bottle!

red bottle

Look – another *one*!

Process: Because the antecedent (“red bottle”) includes the modifier “red”, the property RED is important for the referent of *one* to have.

→ referent of *one* = RED BOTTLE

Pearl & Mis submitted

English anaphoric *one*

Look - a red bottle!

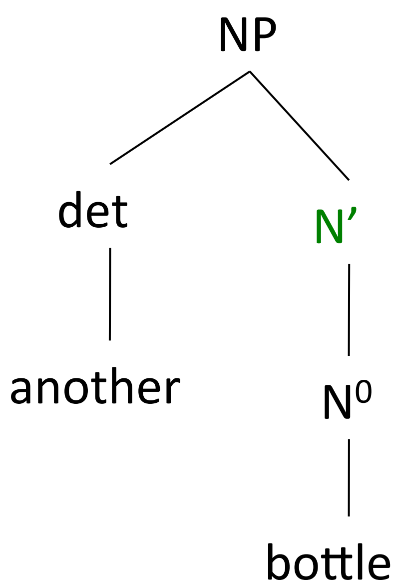
Look – another *one*!

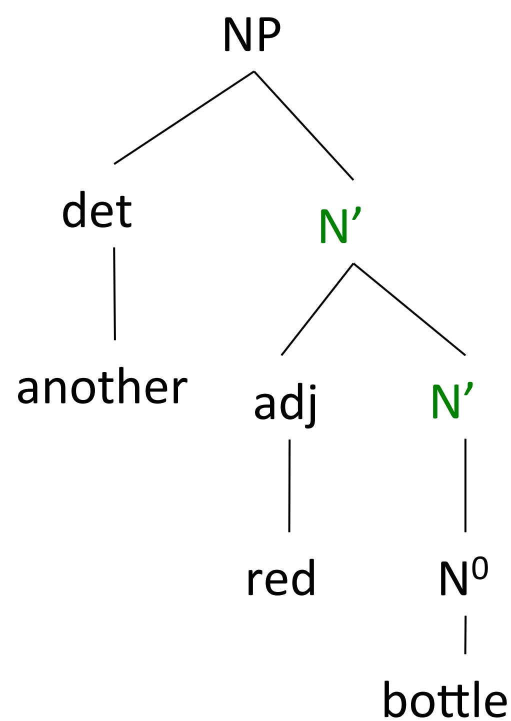
Two steps:
(1) Identify linguistic antecedent
(2) Identify referent (based on linguistic antecedent)

Pearl & Mis submitted

Anaphoric *one*: Syntactic category

Standard linguistic theory (Chomsky 1970, Jackendoff 1977) posits that *one* in these kinds of utterances is a syntactic category smaller than an entire noun phrase (NP), but larger than just a noun (N^0). This category is N'. This category includes strings like "bottle" and "red bottle".





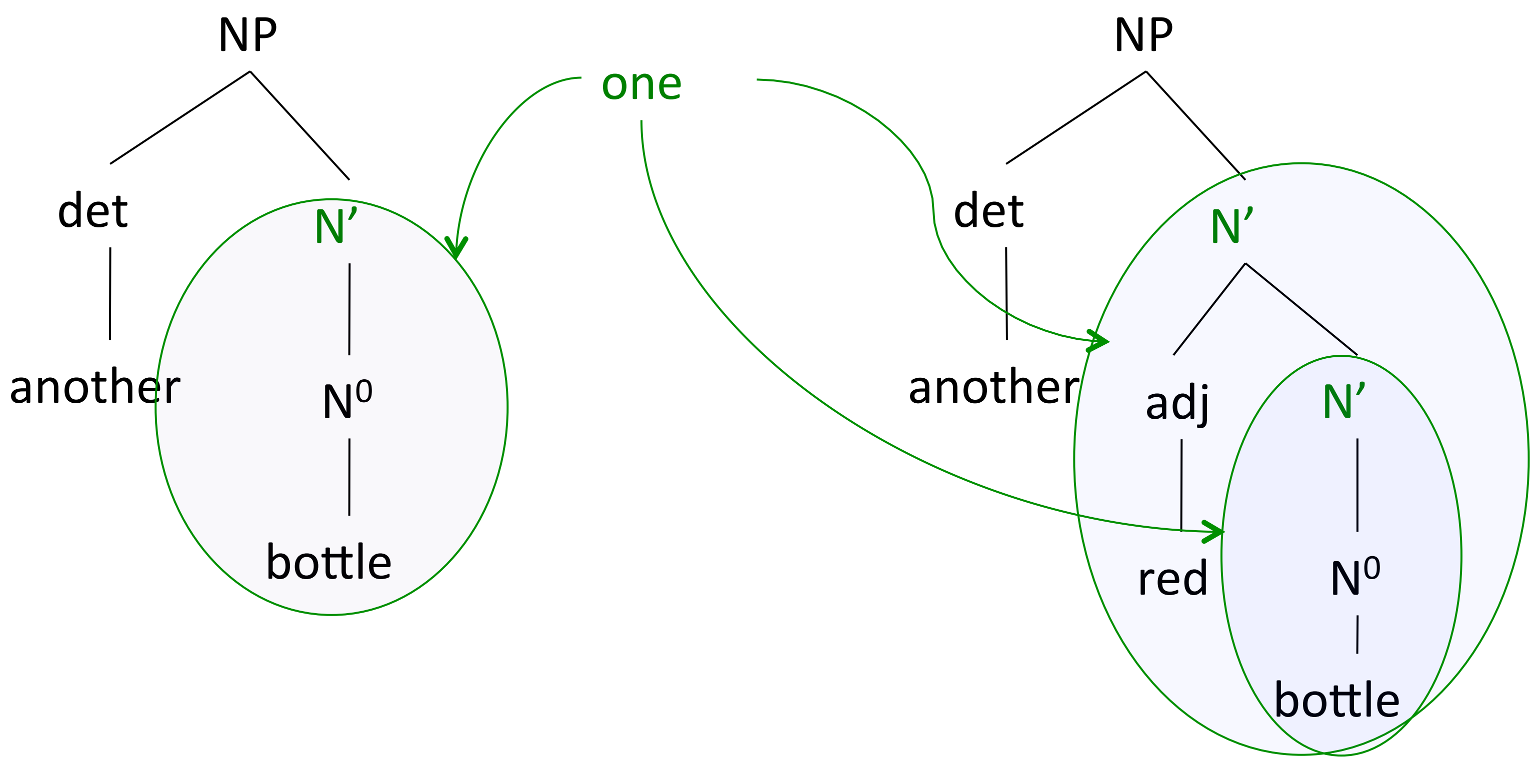
$[_{NP}$ another $[_{N'}$ $[_{N0}$ bottle$]]]$

$[_{NP}$ another $[_{N'}$ red $[_{N'}$ $[_{N0}$ bottle$]]]]$

Pearl & Mis submitted

Anaphoric *one*: Syntactic category

Standard linguistic theory (Chomsky 1970, Jackendoff 1977) posits that *one* in these kinds of utterances is a syntactic category smaller than an entire noun phrase (NP), but larger than just a noun (N^0). This category is N'. This category includes strings like "bottle" and "red bottle".



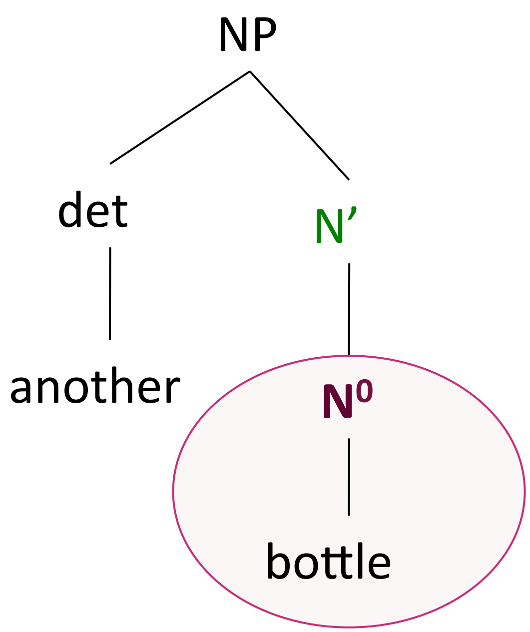
$[_{NP}$ another $[_{N'}$ $[_{N0}$ bottle]]]

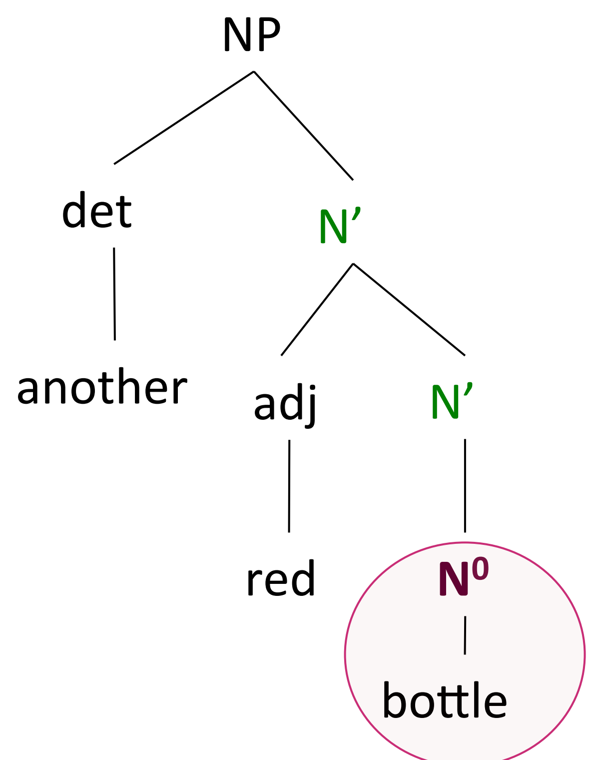
$[_{NP}$ another $[_{N'}$ red $[_{N'}$ $[_{N0}$ bottle]]]]

Pearl & Mis submitted

Anaphoric *one*: Syntactic category

Importantly, *one* is not N^0. If it was, it could only have strings like “bottle” as its antecedent, and could never have strings like “red bottle” as its antecedent.





$[_{NP}$ another $[_{N'}$ $[_{N0}$ bottle]]]

$[_{NP}$ another $[_{N'}$ red $[_{N'}$ $[_{N0}$ bottle]]]]

Pearl & Mis submitted

Anaphoric *one*: Interpretations based on syntactic category

If *one* was N^0, we would not be able to have the "red bottle" interpretation:

"Look – a red bottle!  Look – another one!"

Because *one*'s antecedent could only be "bottle", we would have to interpret the second part as "Look - another *bottle*!"

Since *one*'s antecedent can be "red bottle", and "red bottle" cannot be N^0, *one* must not be N^0 (in this context at least).

Pearl & Mis submitted

Anaphoric *one*: Adult knowledge

"Look – a red bottle! Look – another one!"

≈ "Look – a red bottle! Look – another red bottle!"

Target knowledge state:

Syntactic knowledge: category N'

Referential knowledge: mentioned property ("red") is included in the linguistic antecedent (antecedent = "red bottle"), so referent has property.

Pearl & Mis submitted

Anaphoric *one*: Adult knowledge

"Look – a red bottle! Look – another one!"

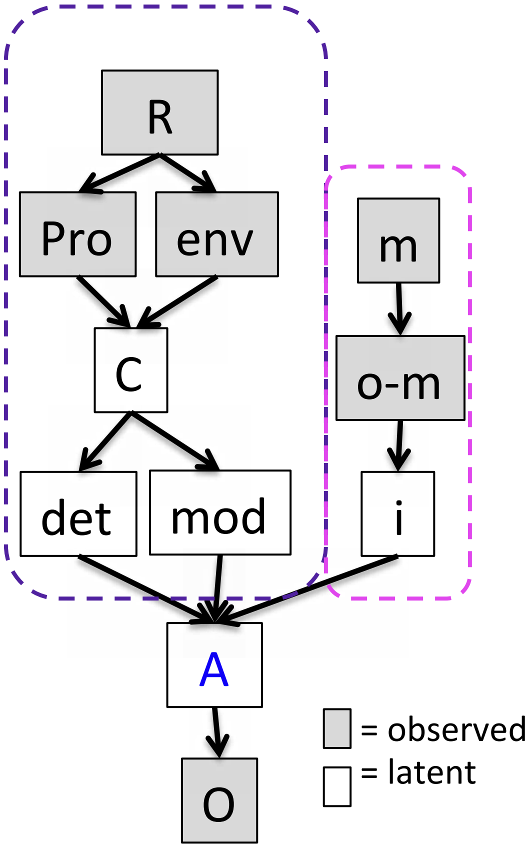
≈ "Look – a red bottle! Look – another red bottle!"

Target behavior state (based on target knowledge state):

In this scenario, adults expect to see another red bottle – not just another bottle. So, they will look for a second red bottle.

Pearl & Mis submitted

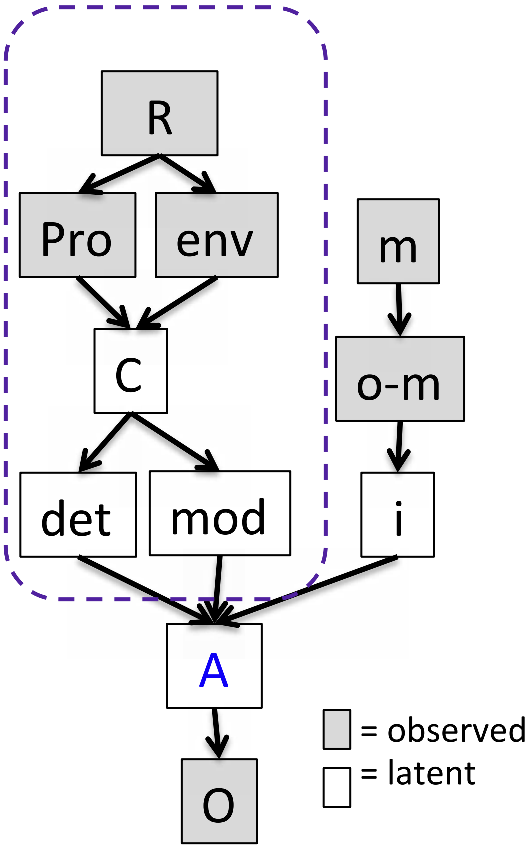
Understanding a referential expression



Includes both syntactic and referential information, since both are used to determine the linguistic antecedent.

Pearl & Mis submitted

Understanding a referential expression



“Look, a red bottle! Look – another one!”

Syntactic information

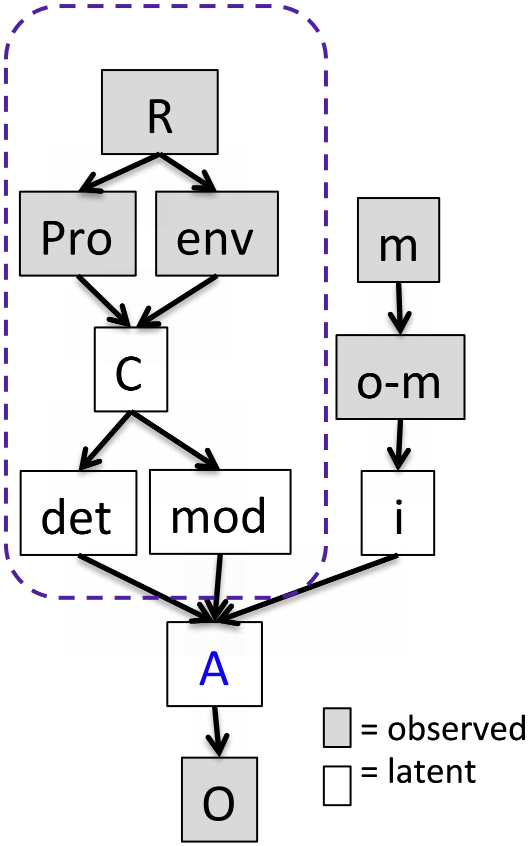
R = referential expression used
ex: “another one”

Pro = pronoun used in referential expression
ex: “one”

env = smaller than NP?
ex: yes

Pearl & Mis submitted

Understanding a referential expression



“Look, a red bottle! Look – another one!”

Syntactic information

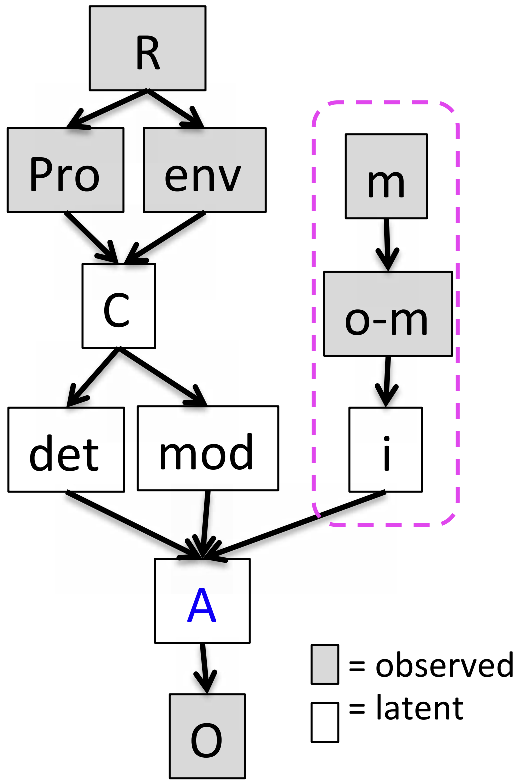
C = syntactic category of pronoun used (= syntactic category of linguistic antecedent)
  ex: N’

det = antecedent includes determiner?
  ex: no

mod = antecedent includes modifier?
  ex: yes

Pearl & Mis submitted

Understanding a referential expression



"Look, a red bottle! Look – another one!"

Referential information

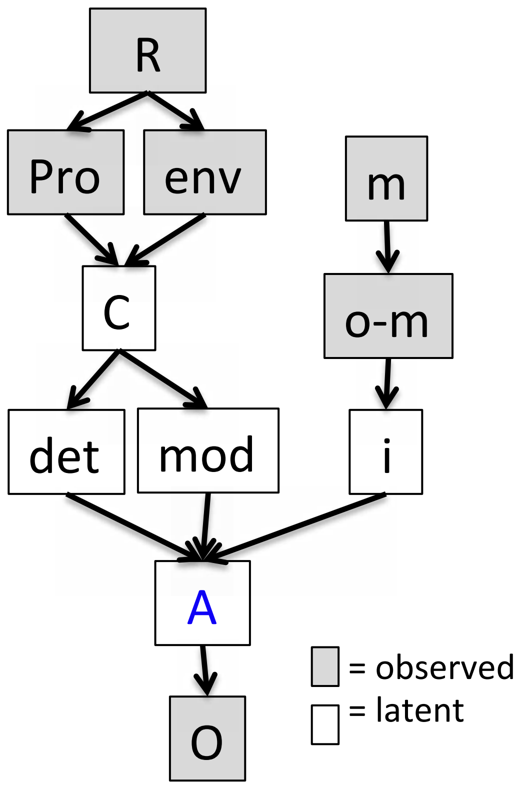
m = property mentioned in previous linguistic context
ex: yes

o-m = referent (object) in current context has mentioned property
ex: yes

i = mentioned property is included in antecedent?
ex: yes

Pearl & Mis submitted

Understanding a referential expression



"Look, a red bottle! Look – another one!"

A = antecedent
ex: "red bottle"
(depends on both syntactic information of *det* and *mod*, and referential information from *i*.)

O = intended object (learner can often observe this)
ex: RED BOTTLE

Pearl & Mis submitted

Anaphoric *one*: Children's knowledge

Lidz, Waxman, & Freedman (2003) [**LWF**]
investigated 18-month-old behavior in this scenario.

"Look – a red bottle!" 

"Now look…"

Control:
"What do you see now?"

Anaphoric:
"Do you see another one?"

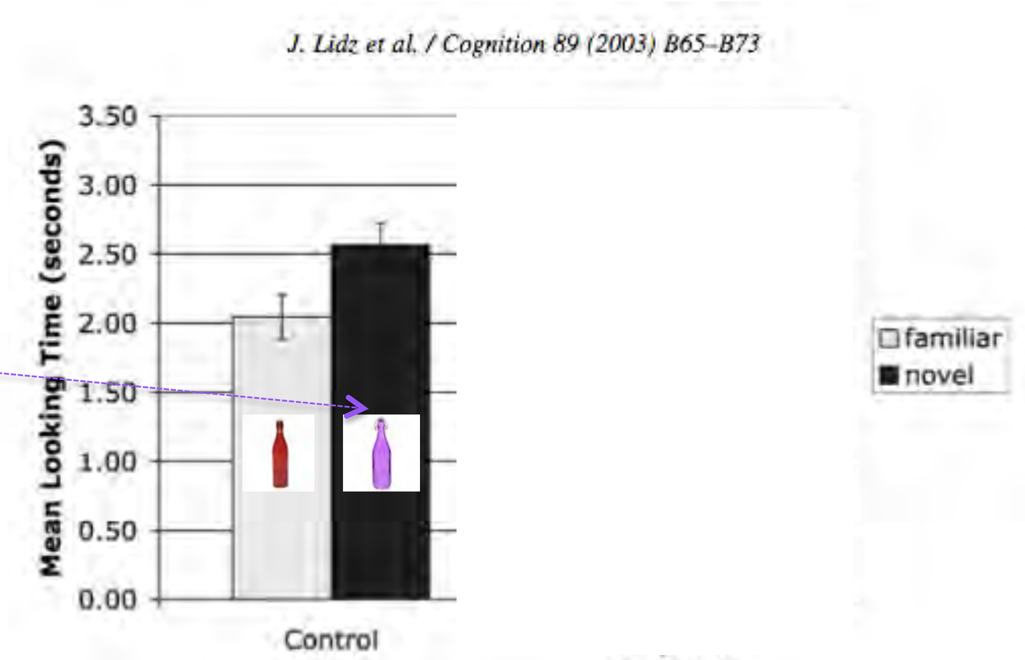
Pearl & Mis submitted

Anaphoric *one*: Children's knowledge

Lidz, Waxman, & Freedman (2003) [**LWF**]
investigated 18-month-old behavior in this scenario.

"Look – a red bottle!" 

"Now look…"



Control:

"What do you see now?"

Baseline novelty preference:

[~2.0s  vs. ~2.5s ]

Anaphoric:

"Do you see another one?"

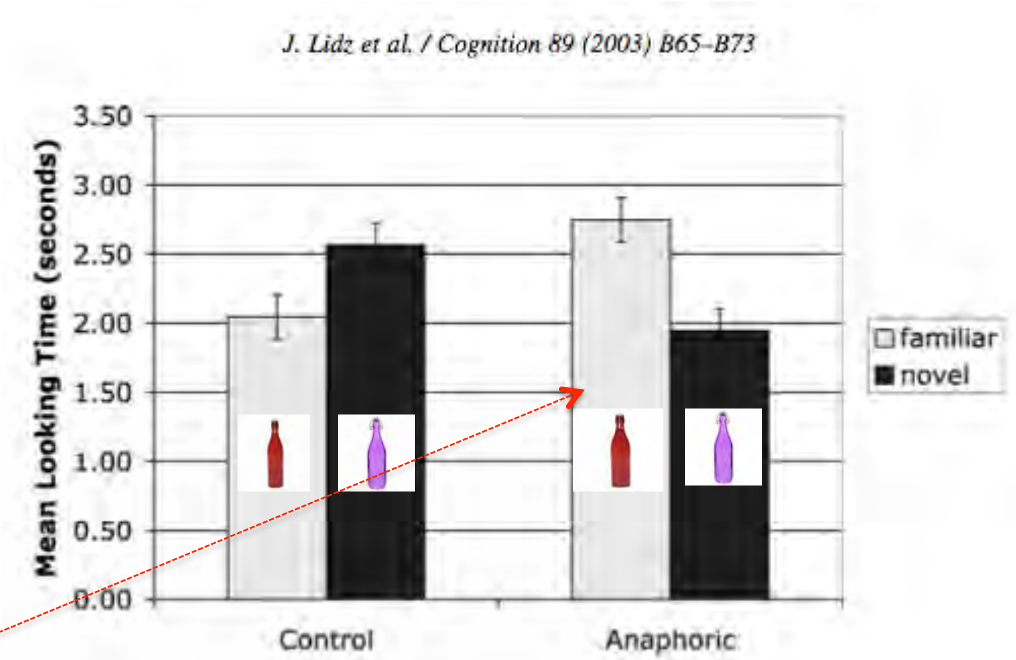
Pearl & Mis submitted

Anaphoric *one*: Children's knowledge

Lidz, Waxman, & Freedman (2003) [**LWF**]
investigated 18-month-old behavior in this scenario.

"Look – a red bottle!"

"Now look…"



Control:
"What do you see now?"
Baseline novelty preference

Anaphoric:
"Do you see another one?"
[~2.75  vs. ~1.95s ]
Adjusted familiarity preference

Pearl & Mis submitted

Anaphoric *one*: Children's knowledge

Lidz, Waxman, & Freedman (2003) [**LWF**]
investigated 18-month-old behavior in this scenario.

"Look – a red bottle!" 

"Now look…"

Noun:
"Do you see another bottle?"

Adjective-noun:
"Do you see another red bottle?"

Pearl & Mis submitted

Anaphoric *one*: Children's knowledge

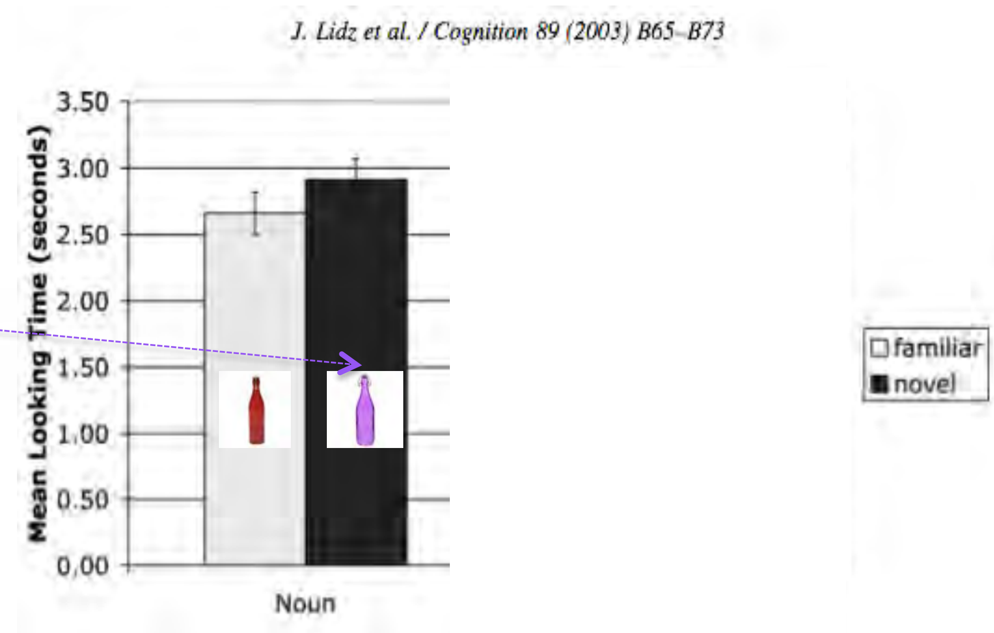
Lidz, Waxman, & Freedman (2003) [**LWF**]
investigated 18-month-old behavior in this scenario.

"Look – a red bottle!" 

"Now look…"

Noun:
"Do you see another bottle?"
Baseline novelty preference:
[~2.65s  vs. ~2.95s ]

Adjective-noun:
"Do you see another red bottle?"



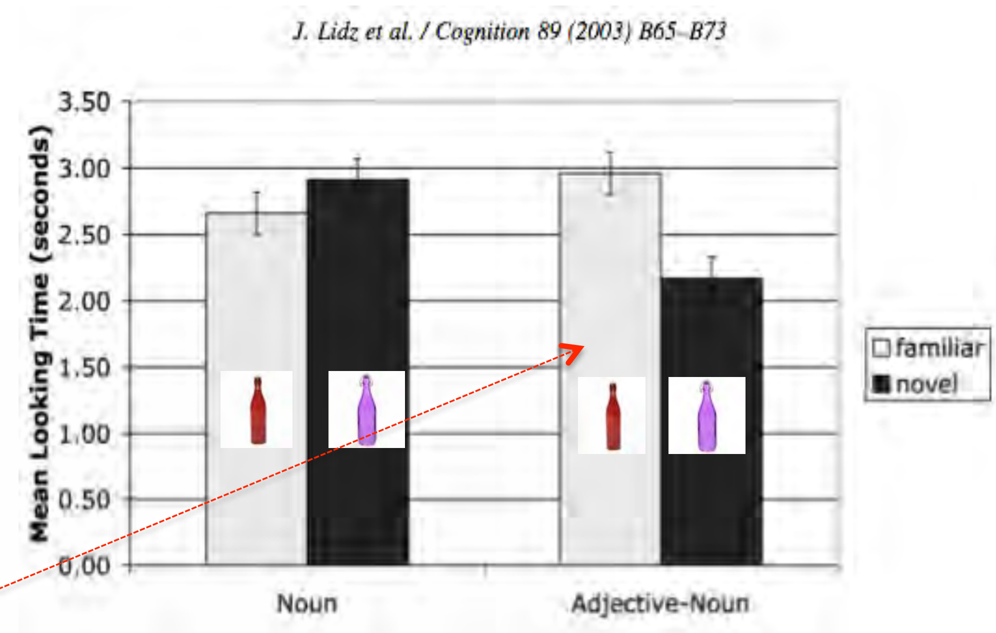
Pearl & Mis submitted

Anaphoric *one*: Children's knowledge

Lidz, Waxman, & Freedman (2003) [**LWF**]
investigated 18-month-old behavior in this scenario.

"Look – a red bottle!" 

"Now look…"



Noun:
"Do you see another bottle?"
Baseline novelty preference

Adjective-noun:
"Do you see another red bottle?"
[~3.0s  vs. ~2.1s ]
Adjusted familiarity preference

Pearl & Mis submitted

Anaphoric *one*: Children's knowledge

Lidz, Waxman, & Freedman (2003) [**LWF**]
investigated 18-month-old behavior in this scenario.

"Look – a red bottle!"

"Now look…"

Control/Noun:

"What do you see now?"

"Do you see another bottle?"

Baseline novelty preference

Average probability of looking to familiar bottle: 0.459

Anaphoric/Adjective-Noun:

"Do you see another one?"

"Do you see another red bottle?"

Adjusted familiarity preference

Average probability of looking to familiar bottle: 0.587

Pearl & Mis submitted

Anaphoric *one*: Children's knowledge

Lidz, Waxman, & Freedman (2003) [**LWF**]
investigated 18-month-old behavior in this scenario.

LWF interpretation:

Given 18-month-olds' baseline novelty preference and adjusted familiarity preference, preference for RED BOTTLE means the preferred antecedent is "red bottle".

Pearl & Mis submitted

Anaphoric *one*: Children's knowledge

Lidz, Waxman, & Freedman (2003) [**LWF**]
investigated 18-month-old behavior in this scenario.

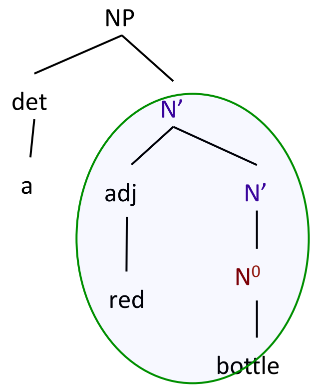
LWF interpretation:

Given 18-month-olds' baseline novelty preference and adjusted familiarity preference, preference for RED BOTTLE means the preferred antecedent is "red bottle".

LWF conclusion about 18-month-old knowledge state:

(1) syntactic category of *one* = N'

(2) linguistic antecedent when modifier is present (i.e., property is mentioned) includes modifier (e.g., "red") = referent has modifier property



one =

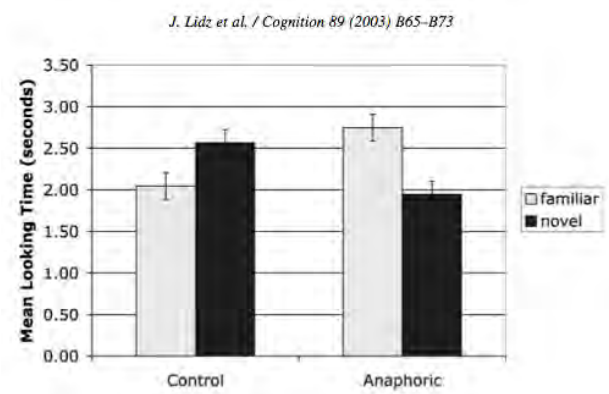
Pearl & Mis submitted

Defining the learning task

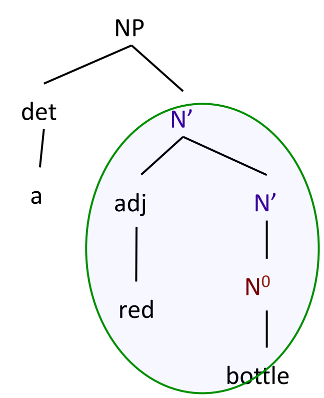
Given a specific initial state, a learner must use the data intake to reach the target state by the end of the learning period.

Target state:

18-month-old behavior = Adjusted familiarity preference when modifier is present in potential antecedent and anaphoric *one* is used (LWF)



Knowledge = *one* is N', the antecedent contains the modifier ("red bottle")



one =

Pearl & Mis submitted

Defining the learning task

Given a specific initial state, a learner must use the data intake to reach the target state by the end of the learning period.

Learning period:

Completed by 18 months (LWF)

Starts?

Pearl & Mis submitted

Defining the learning task

Given a specific initial state, a learner must use the data intake to reach the target state by the end of the learning period.

Learning period:

Completed by 18 months (LWF)

Starts?

Pearl & Lidz 2009 estimate, based on Booth & Waxman (2003):

Children could start learning *one*'s representation as early as 14 months, when they have some grammatical category knowledge.

Pearl & Mis submitted

Defining the learning task

Given a specific initial state, a learner must use the data intake to reach the target state by the end of the learning period.

Learning period:

Completed by 18 months (LWF)

Starts at 14 months

Total time period: 4 months (between 14 – 18 months)

Pearl & Mis submitted

Defining the learning task

Given a specific initial state, a learner must use the data intake to reach the target state by the end of the learning period.

Data intake:

All input data deemed informative.

How do we know what counts as informative?

This is defined by biases in the initial state.

Pearl & Mis submitted

The data intake: Different data types

Direct positive evidence: Unambiguous

Unambiguous *one* (DirUnamb) data:

"Look – a red bottle!

Hmmm - there doesn't seem to be another one here, though."

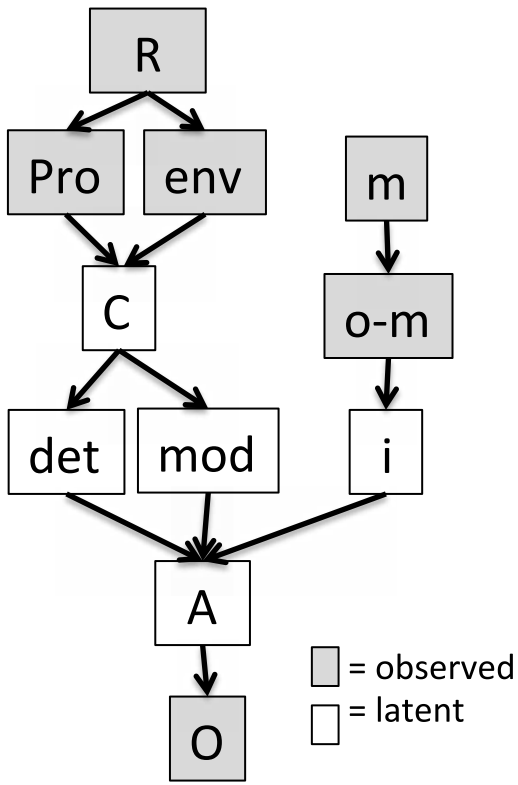
one's referent = BOTTLE? If so, *one*'s antecedent = "bottle".

But it's strange to claim there's not another *bottle* here.

So, *one*'s referent must be RED BOTTLE, and *one*'s antecedent = $[_{N'}$ red$[_{N'}[_{N0}$ bottle$]]]$.

Pearl & Mis submitted

DirUnamb data



"Look, a red bottle! Hmm – there doesn't seem be another one here, though!"

R = "another one"
Pro = "one"
env = <NP

m = yes
o-m = yes

C = N'
det = no
mod = yes

i = yes

A = "red bottle"
O = RED BOTTLE

Pearl & Mis submitted

The data intake: Different data types

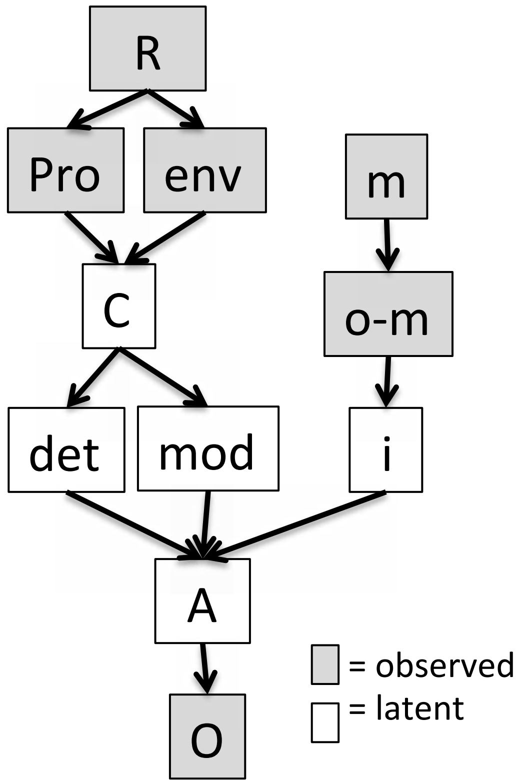
Direct positive evidence: Ambiguous

Syntactically ambiguous (DirSynAmb) data:
"Look – a bottle! Oh, look – another one."

one's referent = BOTTLE
one's antecedent = $[_{N'}[_{N0}$ bottle]] or $[_{N0}$ bottle]?

Pearl & Mis submitted

DirSynAmb data



"Look – a bottle! Oh, look – another one!"

R = "another one"
Pro = "one"
env = <NP
m = no
o-m = N/A

C = N' or N^0?
det = no
mod = no
i = N/A

A = "bottle"
O = BOTTLE

Pearl & Mis submitted

The data intake: Different data types

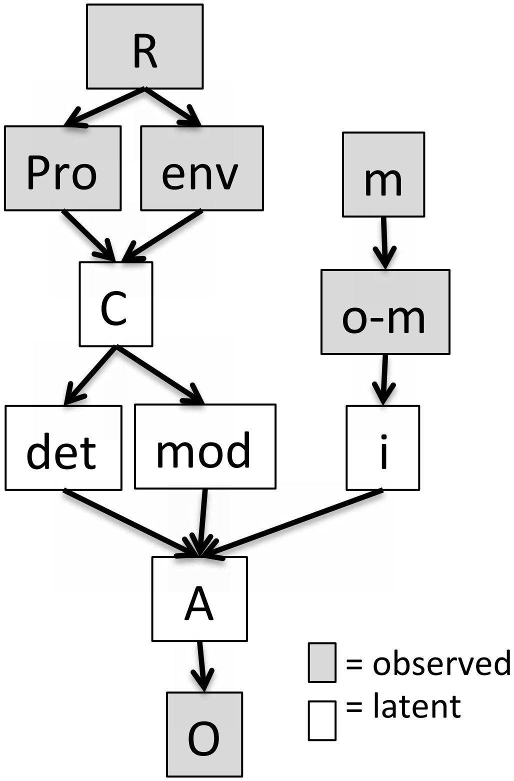
Direct positive evidence: Ambiguous

Referentially and syntactically ambiguous (DirRefSynAmb) data:
"Look – a red bottle! Oh, look – another one."


one's referent = RED BOTTLE or BOTTLE?
one's antecedent = $[_{N'}$ red$[_{N'}[_{N0}$ bottle$]]]$ or $[_{N'}[_{N0}$ bottle$]]$ or $[_{N0}$ bottle$]$?

Pearl & Mis submitted

DirRefSynAmb data



"Look – a red bottle! Oh, look – another one."

R = "another one"

Pro = "one" m = yes

env = <NP o-m = yes

C = N' or N^0?

det = no

mod = yes or no? **i = yes or no?**

A = "red bottle" or "bottle"?

O = RED BOTTLE

Pearl & Mis submitted

The data intake: Different data types

Indirect positive evidence: Unambiguous

Observation: Other words in the language can also be used anaphorically:

him, her, it, …

Look at the cute penguin. I want to hug it.

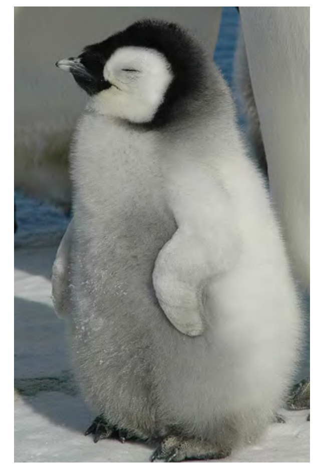
[$_{NP}$ the [$_{N'}$ cute [$_{N'}$ [$_{N0}$ penguin]]]] $\leftarrow$ [$_{NP}$ it]

CUTE PENGUIN

Look! A cute penguin. I want one.

[$_{NP}$ a [$_{N'}$ cute [$_{N'}$ [$_{N0}$ penguin]]]] $\leftarrow$ [$_{NP}$ one]

CUTE PENGUIN



Pearl & Mis submitted

The data intake: Different data types

Indirect positive evidence: Unambiguous

Syntactic information:
So the antecedent should be an NP, which includes the modifier.

Syntactic information:
Pronoun is NP

Look at the cute penguin. I want to hug it.

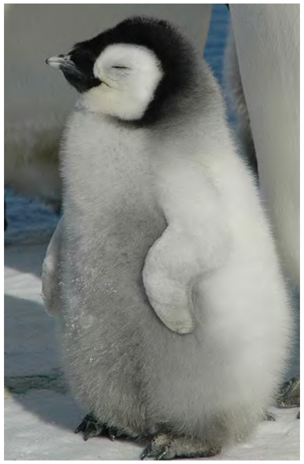
$[_{NP}$ the $[_{N'}$ cute $[_{N'}$ $[_{N0}$ penguin$]]]]$ ← $[_{NP}$ it$]$

CUTE PENGUIN

Look! A cute penguin. I want one.

$[_{NP}$ a $[_{N'}$ cute $[_{N'}$ $[_{N0}$ penguin$]]]]$ ← $[_{NP}$ one$]$

CUTE PENGUIN



Pearl & Mis submitted

The data intake: Different data types

Indirect positive evidence: Unambiguous

This indirect positive evidence coming from other pronoun data (IndirUnamb) is unambiguous with respect to syntactic category and referent.

Look at the cute penguin. I want to hug it.

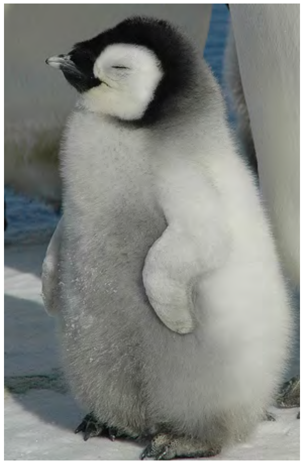
$[_{NP}$ the $[_{N'}$ cute $[_{N'}$ $[_{N0}$ penguin$]]]]$ ← $[_{NP}$ it$]$

CUTE PENGUIN

Look! A cute penguin. I want one.

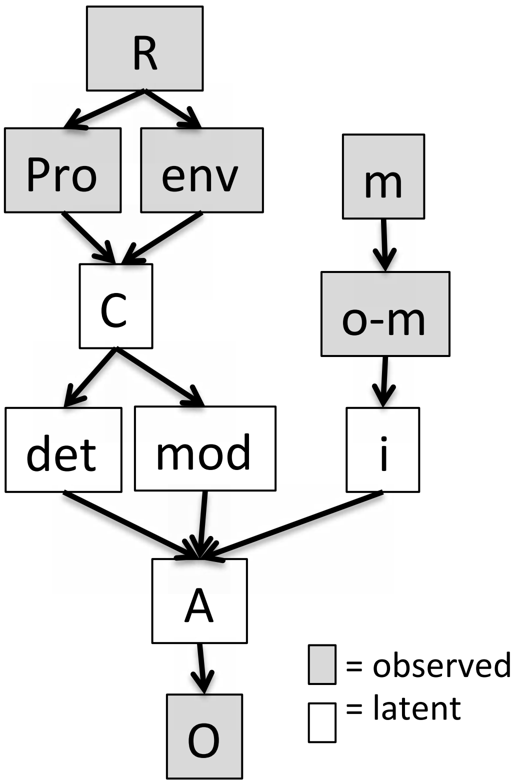
$[_{NP}$ a $[_{N'}$ cute $[_{N'}$ $[_{N0}$ penguin$]]]]$ ← $[_{NP}$ one$]$

CUTE PENGUIN



Pearl & Mis submitted

IndirUnamb data



"Look - a red bottle! I want it."

R = "it"

Pro = "it"

env = NP

m = yes

o-m = yes

C = NP

det = yes

mod = yes

i = yes

A = "a red bottle"

O = RED BOTTLE

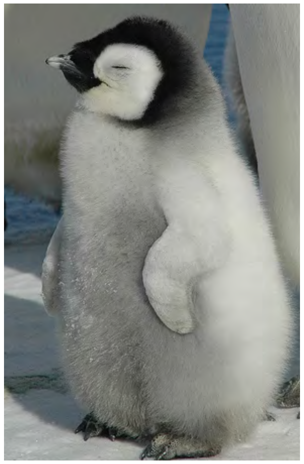
Pearl & Mis submitted

The utility of indirect positive evidence

The IndirUnamb data coming from indirect positive evidence can also be used to determine how often the referent of the anaphoric element has the mentioned property.

Referential information:
Is the referent cute? Yes!

Look at the cute penguin. I want to hug it.



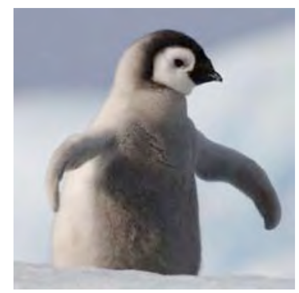
Pearl & Mis submitted

The utility of indirect positive evidence

These data can help bias learner expectations when encountering pronouns that have more than one potential antecedent.

Look! A cute penguin.

There's another one.



How often do the referents contain the mentioned property? [$_{<NP}$ one]

Often

[$_{N'}$ cute [$_{N'}$ [$_{N0}$ penguin]]]

CUTE PENGUIN

Not often

[$_{N'}$ [$_{N0}$ penguin]] or [$_{N0}$ penguin]

PENGUIN

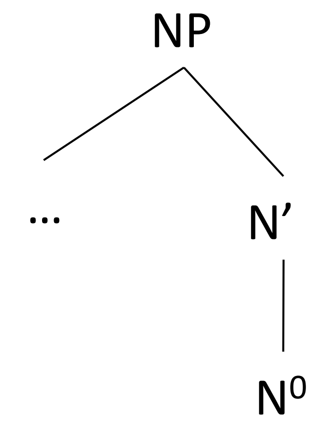
Pearl & Mis submitted

Defining the learning task

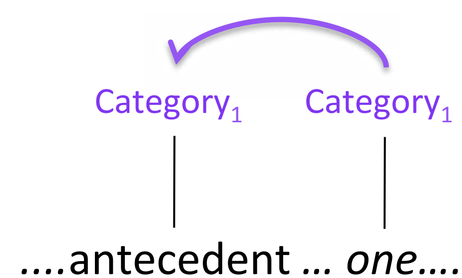
Given a specific initial state, a learner must use the data intake to reach the target state by the end of the learning period.

Initial state:

Knowledge: Syntactic categories exist, in particular N^0, N', and NP.



Knowledge: Anaphoric elements like *one* take linguistic antecedents of the same category.



Pearl & Mis submitted

Learning strategies: Updating the initial state

Initial state:

- Knowledge: Syntactic categories exist, in particular N^0, N', and NP.
- Knowledge: Anaphoric elements like *one* take linguistic antecedents of the same category.

DirUnamb learner (Baker 1978, Hornstein & Lightfoot 1981, Lightfoot 1982, Hamburger & Crain 1984, Crain 1991)

Initial state update:
+ Only direct positive unambiguous data are informative.

Data intake specification:
Informative data = DirUnamb

Previous DirUnamb finding: This learner has almost no data to learn from and fails to learn the target knowledge.

Pearl & Mis submitted

Learning strategies: Updating the initial state

Initial state:

- Knowledge: Syntactic categories exist, in particular N^0, N', and NP.
- Knowledge: Anaphoric elements like *one* take linguistic antecedents of the same category.

DirUnamb + N' learner (Baker 1978)

Initial state update:
+ Only direct positive unambiguous data are informative.
+ *One* is not N^0

Data intake specification:
Informative data = DirUnamb

Previous DirUnamb + N' finding: While there's still little data to learn from, this learner already has the target syntactic knowledge. (If *one* is not N^0 in these contexts, it is N'.)

Pearl & Mis submitted

Learning strategies: Updating the initial state

Initial state:

- ❖ Knowledge: Syntactic categories exist, in particular N^0, N', and NP.
- ❖ Knowledge: Anaphoric elements like *one* take linguistic antecedents of the same category.

DirFiltered learner (Regier & Gahl 2004, Pearl & Lidz 2009)

Initial state update:
+ Direct positive and indirect negative data are informative.
+ Use probabilistic inference
+ Filter out DirSynAmb data

Data intake specification:
Informative data = DirUnamb, DirRefSynAmb

Previous DirFiltered finding: With more data to learn from, this learner learns the target knowledge.

Pearl & Mis submitted

Learning strategies: Updating the initial state

Initial state:

- Knowledge: Syntactic categories exist, in particular N^0, N', and NP.
- Knowledge: Anaphoric elements like *one* take linguistic antecedents of the same category.

Direct Equal Opportunity (DirEO) learner (Pearl & Lidz 2009)

Initial state update:
+ Direct positive and indirect negative data are informative.
+ Use probabilistic inference

Data intake specification:
Informative data = DirUnamb, DirRefSynAmb, DirSynAmb

Previous DirEO finding: This learner does not learn the target knowledge – the DirSynAmb data lead the learner to the wrong syntactic generalization.

Pearl & Mis submitted

Learning strategies: Updating the initial state

Initial state:

- Knowledge: Syntactic categories exist, in particular N^0, N', and NP.
- Knowledge: Anaphoric elements like *one* take linguistic antecedents of the same category.

Indirect evidence from pronouns (IndirPro) learner (Pearl & Mis 2011, 2013, submitted)

Initial state update:
+ Direct positive, indirect negative, and indirect positive data are informative.
+ Indirect positive evidence = other pronoun data
+ Use probabilistic inference

Data intake specification:
Informative data = DirUnamb, DirRefSynAmb, DirSynAmb, IndirUnamb

IndirPro finding: Let's find out…

Pearl & Mis submitted

Data set comparisons

DirUnamb

"Look – a red bottle! Hmmm - there doesn't seem to be another *one* here, though."

Learners:
DirUnamb, DirUnamb + N', DirFiltered, DirEO, IndirPro

DirRefSynAmb

"Look – a red bottle! Oh, look – another *one*!"

Learners:
DirFiltered, DirEO, IndirPro

DirSynAmb

"Look – a bottle! Oh, look – another *one*!"

Learners:
DirEO, IndirPro

IndirUnamb

"Look – a red bottle! I want *one*/*it*."

Learners:
IndirPro

Pearl & Mis submitted

Corpus analysis & learner input

Brown/Eve corpus (CHILDES: MacWhinney 2000): starting at 18 months

17,521 utterances of child-directed speech, 2874 pronoun utterances

[~16.4% pronoun utterances]

Learning period = 4 months (between 14 and 18 months)

Based on estimates of the number of utterances children hear from birth until 18 months (Akhtar et al., 2004), we can calculate the data distribution in their input between 14 and 18 months (~36,500 pronoun utterances total).

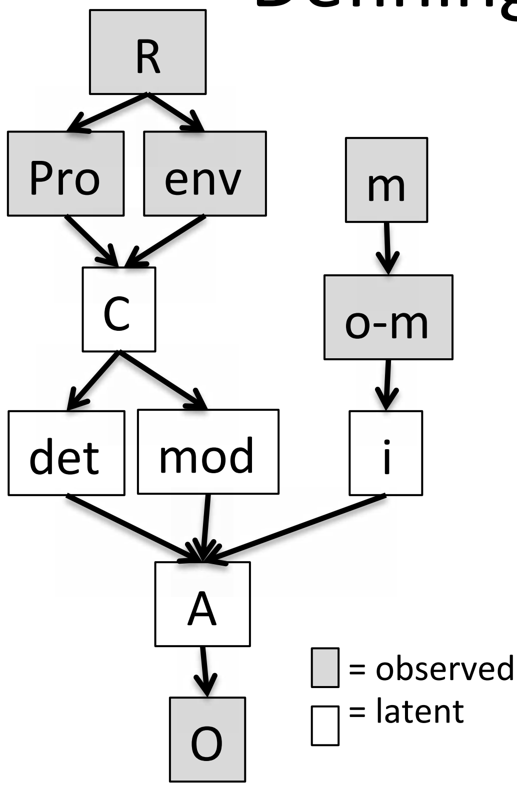
Pearl & Mis submitted

Corpus analysis & learner input

		DirUnamb	DirUnamb + N'	DirFiltered	DirEO	IndirPro
DirUnamb	0.00%	0	0	0	0	0
DirRefSynAmb	0.66%	0	0	242	242	242
DirSynAmb	7.52%	0	0	0	2743	2743
IndirUnamb	8.42%	0	0	0	0	3073
Uninformative	83.4%	36500	36500	36258	33515	30442

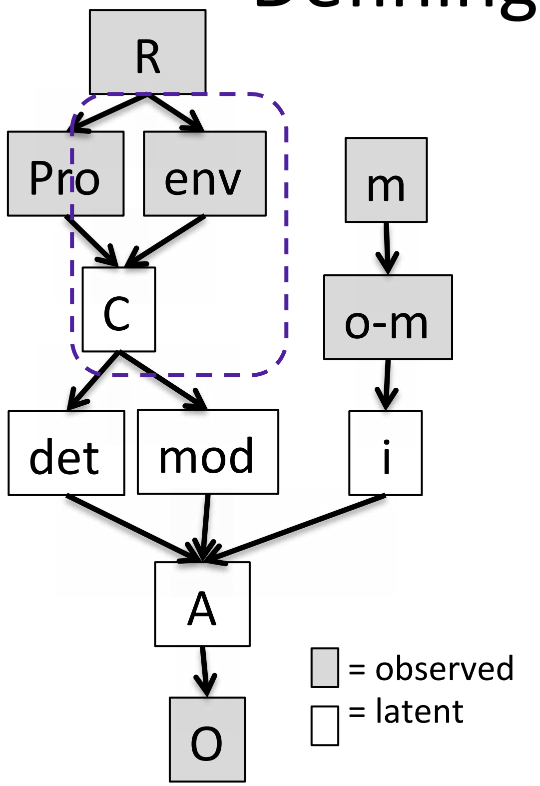
Pearl & Mis submitted

Learning:
Defining target knowledge more formally



Pearl & Mis submitted

Learning:
Defining target knowledge more formally



syntactic knowledge: category of *one*

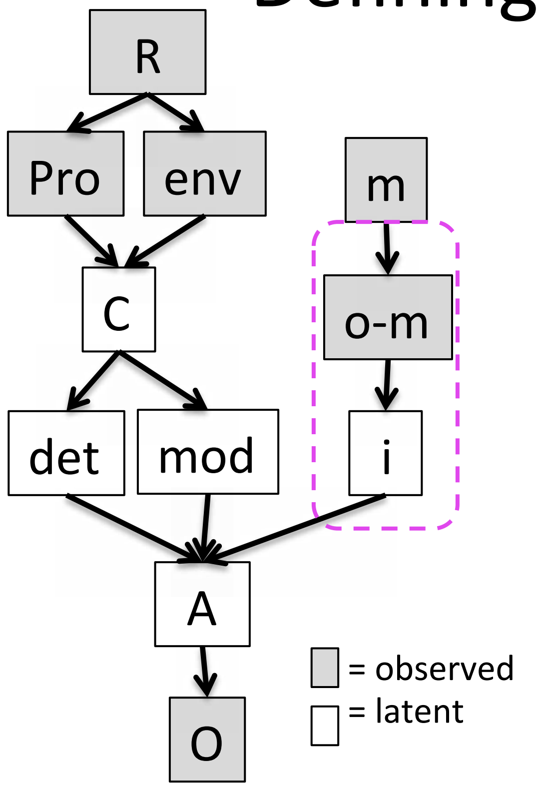
When the syntactic environment indicates the category is smaller than NP (env=<NP), the probability that the syntactic category is N' (C=N'):

$p_{N'}$ = p(C=N' | env=<NP)

Two values: (C=N' or C=N^0)

Pearl & Mis submitted

Learning:
Defining target knowledge more formally



referential knowledge: include property

When an object has the property mentioned in the potential antecedent (o-m=yes), the probability that the property is included in the antecedent (i=yes):

p_{incl} = p(i=yes | o-m=yes)

Two values: (i=yes or i=no)

Pearl & Mis submitted

The online probabilistic learning framework

General form of online update equations for p_x (adapted from Chew 1971):

data seen suggesting x is true

$$p_x = \frac{\alpha + data_x}{\alpha + \beta + totaldata_x}, \alpha = \beta = 1$$

A very weak prior

total informative data seen w.r.t x

After every informative data point encountered:

$$data_x = data_x + \phi_x$$

Incremented by probability that data point suggests x is true

$$totaldata_x = totaldata_x + 1$$

One informative data point seen

Pearl & Mis submitted

Updating $p_{N'}$

$$\phi_{N'} = p(C = N' \mid env = \text{<NP})$$

$$= \frac{p(C = N',\ env = \text{<NP})}{p(env = \text{<NP})}$$

$$= \frac{\sum_{\text{O,A,det,mod,Pro,R,i,o-m,m}} p(C = N',\ env = \text{<NP})}{\sum_{\text{O,A,det,mod,C, Pro,R,i,o-m,m}} p(env = \text{<NP})}$$

Value differs depending on data type:
Direct positive evidence (DirUnamb, DirRefSynAmb, DirSynAmb)
Indirect positive evidence (IndirUnamb)

Pearl & Mis submitted

Updating $p_{N'}$

	Example	$\phi_{N'}$	Intuition
DirUnamb	"…red bottle…don't see another one…"	1	Category definitely N'

Pearl & Mis submitted

Updating $p_{N'}$

	Example	$\phi_{N'}$	Intuition
DirUnamb	"…red bottle…don't see another one…"	1	Category definitely N'
IndirUnamb	"…red bottle… want it…"	N/A	Not informative for $p_{N'}$

Pearl & Mis submitted

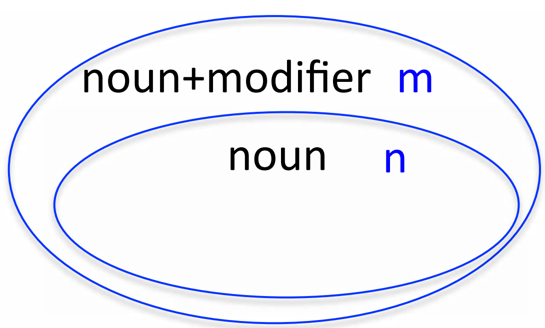
Updating $p_{N'}$

	Example	$\phi_{N'}$	Intuition
DirUnamb	"…red bottle…don't see another one…"	1	Category definitely N'
DirRefSynAmb	"…red bottle…see another one…"	$\frac{rep_1 + rep_2}{rep_1 + rep_2 + rep_3}$	Probability category is N'
IndirUnamb	"…red bottle… want it…"	N/A	Not informative for $p_{N'}$

"red bottle"

$$rep_1 = p_{N'} * \frac{m}{m+n} * p_{incl}$$

Category = N', choose N' with modifier, property is included



N' uses

Pearl & Mis submitted

Updating $p_{N'}$

	Example	$\phi_{N'}$	Intuition
DirUnamb	"…red bottle…don't see another one…"	1	Category definitely N'
DirRefSynAmb	"…red bottle…see another one…"	$\frac{rep_1 + rep_2}{rep_1 + rep_2 + rep_3}$	Probability category is N'
IndirUnamb	"…red bottle… want it…"	N/A	Not informative for $p_{N'}$

"red bottle"

$$rep_1 = p_{N'} * \frac{m}{m+n} * p_{incl}$$

Category = N', choose N' with modifier, property is included

$$rep_2 = p_{N'} * \frac{n}{m+n} * (1 - p_{incl}) * \frac{1}{s}$$

Category = N', choose N' without modifier, property is not included, object has property by chance

"bottle"

$$rep_3 = (1 - p_{N'}) * (1 - p_{incl}) * \frac{1}{s}$$

Category = N^0, property is not included, object has property by chance

Pearl & Mis submitted

Updating $p_{N'}$

	Example	$\phi_{N'}$	Intuition
DirUnamb	"…red bottle…don't see another one…"	1	Category definitely N'
DirRefSynAmb	"…red bottle…see another one…"	$\frac{rep_1 + rep_2}{rep_1 + rep_2 + rep_3}$	Probability category is N'
DirSynAmb	"…bottle…see another one…"	$\frac{rep_4}{rep_4 + rep_5}$	Probability category is N'
IndirUnamb	"…red bottle… want it…"	N/A	Not informative for $p_{N'}$

$$rep_4 = p_{N'} * \frac{n}{m+n}$$

Category = N', choose N' without modifier

"bottle"

$$rep_5 = 1 - p_{N'}$$

Category = N^0

Pearl & Mis submitted

Updating p_{incl}

$$\phi_{incl} = p(i = yes \mid \text{o-m} = yes)$$

$$= \frac{p(i = yes,\ \text{o-m} = yes)}{p(\text{o-m} = yes)}$$

$$= \frac{\sum_{\text{O,A,det,mod,C,Pro,env,R,m}} p(i = yes,\ \text{o-m} = yes)}{\sum_{\text{O,A,det,mod,C, Pro,env,R,i,m}} p(\text{o-m} = yes)}$$

Value differs depending on data type:

Direct positive evidence (DirUnamb, DirRefSynAmb, DirSynAmb)

Indirect positive evidence (IndirUnamb)

Pearl & Mis submitted

Updating p_{incl}

	Example	ϕ_{incl}	Intuition
DirUnamb	"…red bottle…don't see another one…"	1	Property definitely included

Pearl & Mis submitted

Updating p_{incl}

	Example	ϕ_{incl}	Intuition
DirUnamb	"…red bottle…don't see another one…"	1	Property definitely included
IndirUnamb	"…red bottle… want it…"	1	Property definitely included

Pearl & Mis submitted

Updating p_{incl}

	Example	ϕ_{incl}	Intuition
DirUnamb	"…red bottle…don't see another one…"	1	Property definitely included
DirSynAmb	"…bottle…see another one…"	N/A	Not informative for p_{incl}
IndirUnamb	"…red bottle… want it…"	1	Property definitely included

Pearl & Mis submitted

Updating p_{incl}

	Example	ϕ_{incl}	Intuition
DirUnamb	"…red bottle…don't see another one…"	1	Property definitely included
DirRefSynAmb	"…red bottle…see another one…"	$\frac{rep_1}{rep_1 + rep_2 + rep_3}$	Probability property included
DirSynAmb	"…bottle…see another one…"	N/A	Not informative for p_{incl}
IndirUnamb	"…red bottle… want it…"	1	Property definitely included

"red bottle"

$$rep_1 = p_{N'} * \frac{m}{m+n} * p_{incl}$$

Category = N', choose N' with modifier, property is included

$$rep_2 = p_{N'} * \frac{n}{m+n} * (1 - p_{incl}) * \frac{1}{s}$$

Category = N', choose N' without modifier, property is not included, object has property by chance

"bottle"

$$rep_3 = (1 - p_{N'}) * (1 - p_{incl}) * \frac{1}{s}$$

Category = N^0, property is not included, object has property by chance

Pearl & Mis submitted

Example updates

Start with $p_{N'} = p_{incl} = 0.50$, m = 1, n = 2.9, s = 10

One DirUnamb data point: $p_{N'} = 0.67$, $p_{incl} = 0.67$

One DirRefSynAmb data point: $p_{N'} = 0.59$, $p_{incl} = 0.53$

One DirSynAmb data point: $p_{N'} = 0.48$, $p_{incl} = 0.50$

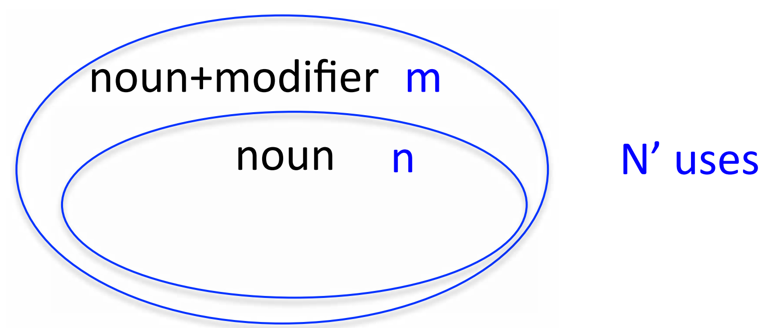
One IndirUnamb data point: $p_{N'} = 0.50$, $p_{incl} = 0.67$

Pearl & Mis submitted

Learner parameters

Free model parameters:
m and *n* (how often N' phrases include modifiers vs. being noun-only)

m=1, n=2.9 (from CHILDES corpus estimate done by Pearl & Lidz 2009)



Pearl & Mis submitted

Learner parameters

Free model parameters:
s (how many salient properties there are – determines how suspicious a coincidence it is if the referent has the mentioned property)

If there are only a few salient properties, it may not be that surprising. However, if there are many salient properties, it becomes more suspicious that the referent just happens to have the mentioned property.

Child may only be aware of a few salient properties or may consider all known properties (# of adjectives known by 16 months ≈ 49 (MacArthur CDI: Dale & Fenson 1996). Pearl & Mis (2013) explored a range from 2 to 49.

Results reported here for s=10.

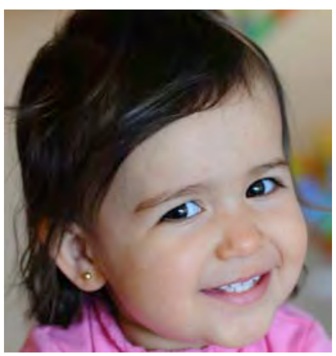
Pearl & Mis submitted

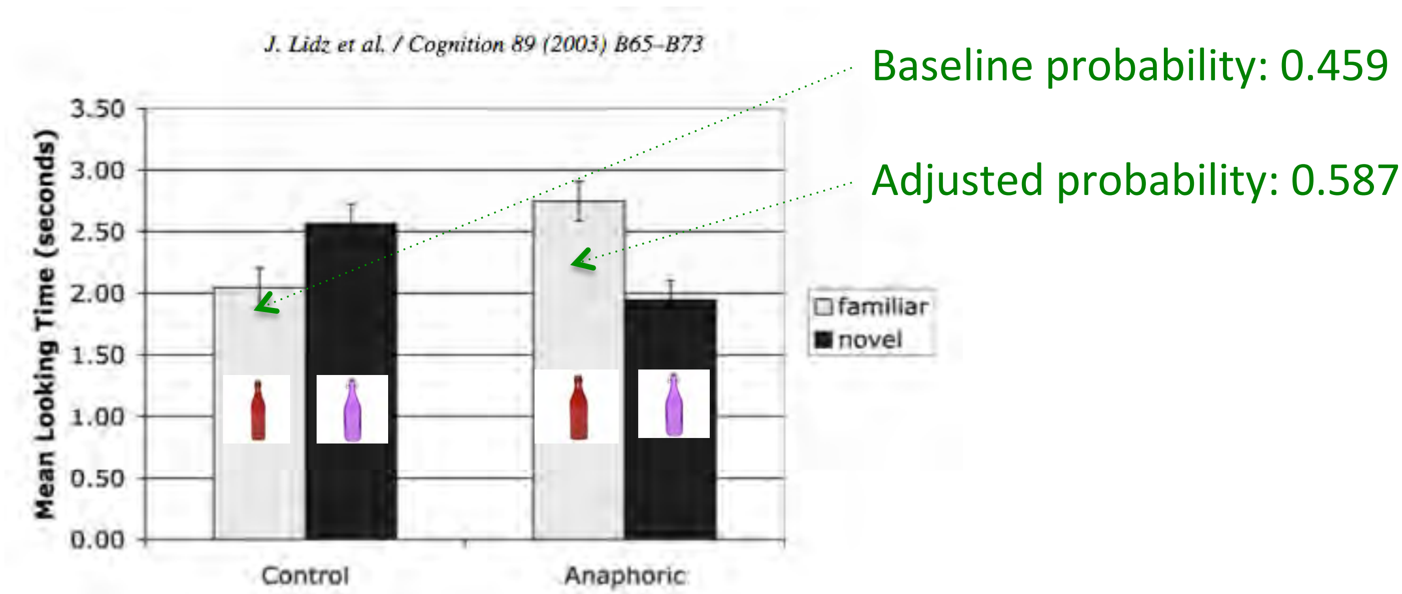
Evaluating learners

Previous investigations have focused on how to learn the target knowledge.

$$p_{N'} = p_{incl} \approx 1.000$$

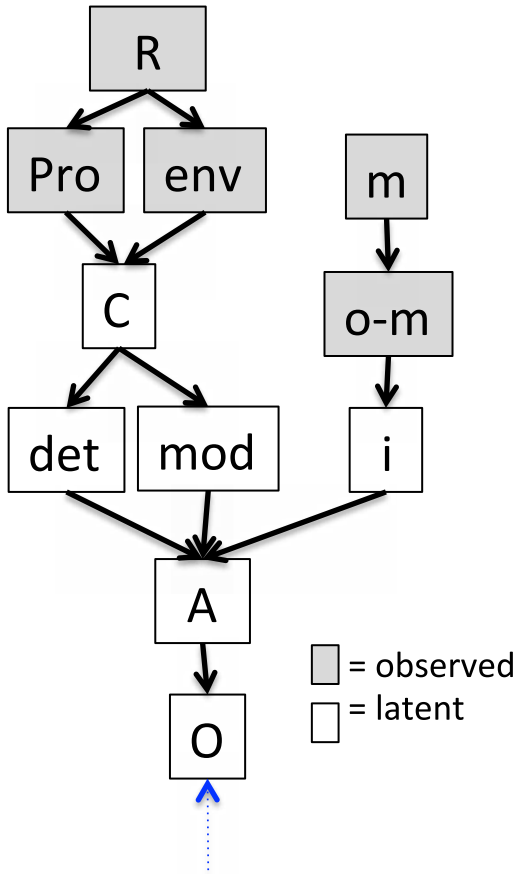
Since we have behavioral data from 18-month-olds, we can also assess how well a learner generates the target behavior (p_{beh}) of looking at the familiar bottle with higher probability when hearing an anaphoric *one* utterance.





Pearl & Mis submitted

Target behavior: p_{beh}

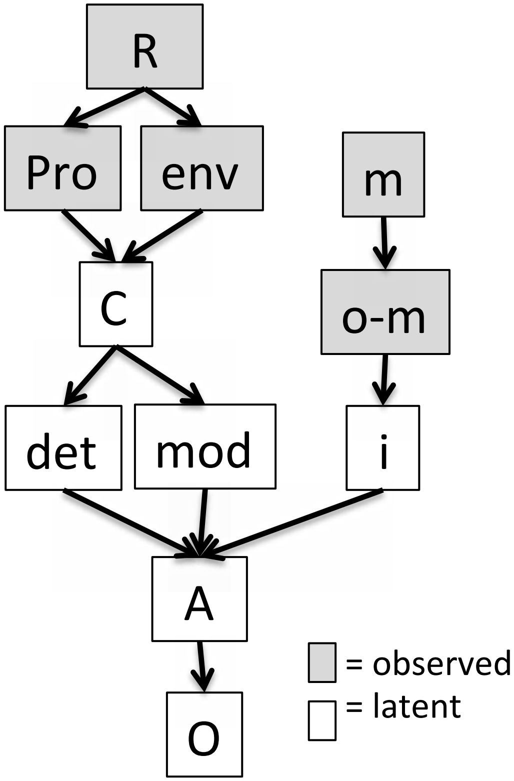


No longer observed –
must be inferred

"Look – a red bottle. Do you see another *one*?"

Pearl & Mis submitted

Target behavior: p_{beh}



O-M = 

$$p_{beh} = p(O = \text{O-M} \mid R = \text{ another one, Pro = one, env = <NP, m=}\mathit{yes}\text{, o-m=}\mathit{yes})$$

$$= \frac{p(O = \text{O-M}, R = \text{ another one, Pro = one, env = <NP, m=}\mathit{yes}\text{, o-m=}\mathit{yes})}{p(R = \text{ another one, Pro = one, env = <NP, m=}\mathit{yes}\text{, o-m=}\mathit{yes})}$$

$$= \frac{\sum_{\text{A,det,mod,C,i}} p(O = \text{O-M}, R = \text{ another one, Pro = one, env = <NP, m=}\mathit{yes}\text{, o-m=}\mathit{yes})}{\sum_{\text{O,A,det,mod,C,i}} p(R = \text{ another one, Pro = one, env = <NP, m=}\mathit{yes}\text{, o-m=}\mathit{yes})}$$

"Look – a red bottle. Do you see another *one*?"

Pearl & Mis submitted

Target behavior: p_{beh}

$$p_{beh} = \frac{rep_{1f} + rep_{2f} + rep_{3f}}{rep_{1f} + rep_{1n} + rep_{2f} + rep_{2n} + rep_{3f} + rep_{3n}}$$

Any outcome where learner looks at (familiar) red bottle

b = baseline preference for looking at familiar bottle = 0.459
a = adjusted preference for looking at familiar bottle when antecedent is "red bottle" = 0.587

$$rep_{1f} = p_{N'} * \frac{m}{m+n} * p_{incl} * a$$

Category = N', antecedent = "red bottle", adjusted familiarity preference

$$rep_{2f} = p_{N'} * \frac{n}{m+n} * (1 - p_{incl}) * b$$

Category = N', antecedent = "bottle", baseline familiarity preference

$$rep_{3f} = (1 - p_{N'}) * (1 - p_{incl}) * b$$

Category = N^0, antecedent = "bottle", baseline familiarity preference

Pearl & Mis submitted

Target behavior: p_{beh}

$$p_{beh} = \frac{rep_{1f} + rep_{2f} + rep_{3f}}{rep_{1f} + rep_{1n} + rep_{2f} + rep_{2n} + rep_{3f} + rep_{3n}}$$

Any outcome where learner looks at (familiar) red bottle

+ Additional outcomes where learner looks at other (novel) bottle

b = baseline preference for looking at familiar bottle = 0.459

a = adjusted preference for looking at familiar bottle when antecedent is "red bottle" = 0.587

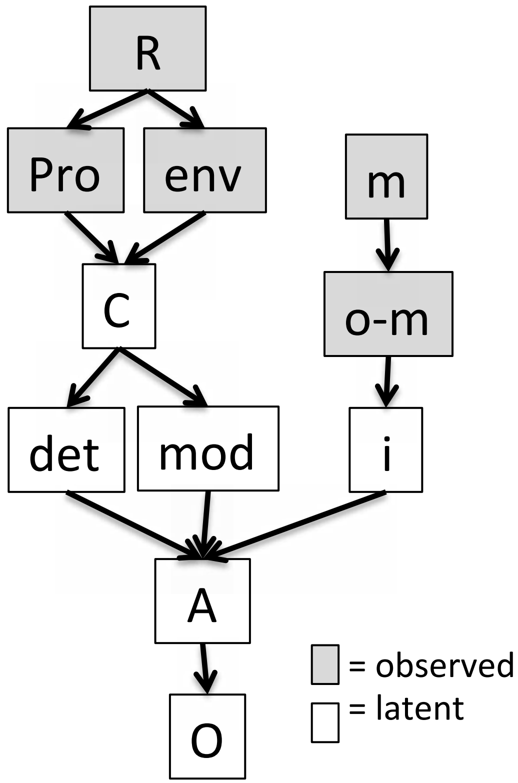
$$rep_{1n} = p_{N'} * \frac{m}{m+n} * p_{incl} * (1-a)$$ Category = N', antecedent = "red bottle", adjusted novelty preference

$$rep_{2n} = p_{N'} * \frac{n}{m+n} * (1-p_{incl}) * (1-b)$$ Category = N', antecedent = "bottle", baseline novelty preference

$$rep_{3n} = (1-p_{N'}) * (1-p_{incl}) * (1-b)$$ Category = N^0, antecedent = "bottle", baseline novelty preference

Pearl & Mis submitted

Context-specific representation: $p_{rep|beh}$



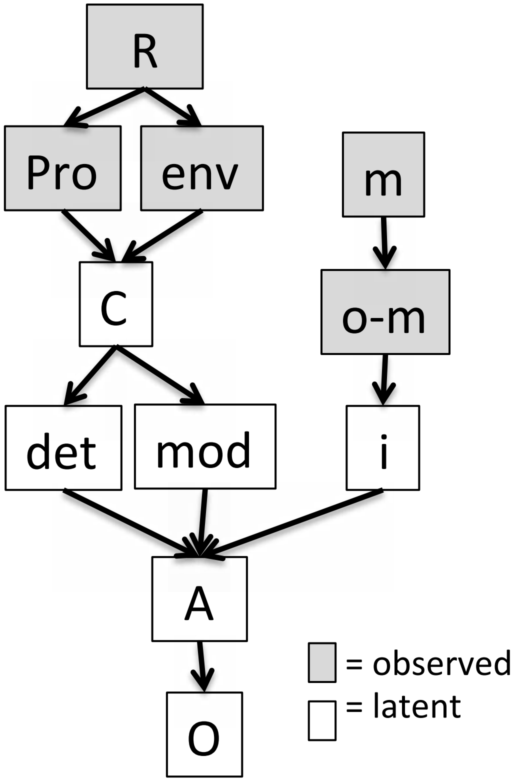
"Look – a red bottle. Do you see another *one*?"

Underlying knowledge check:

When the target behavior is generated, is it being generated because the learner has the target knowledge representation in this context?

Pearl & Mis submitted

Context-specific representation: $p_{rep|beh}$



A = "red bottle"

"Look – a red bottle. Do you see another *one*?"

O-M = 

$$p_{rep|beh} = p(A = \text{red bottle}, i = yes, \text{det} = no, \text{mod} = yes, C = N' \mid R = \text{ another one, Pro = one, env = <NP, m=}yes\text{, o-m=}yes, O = \text{O-M})$$

$$= \frac{p(A = \text{red bottle}, i = yes, \text{det} = no, \text{mod} = yes, C = N', R = \text{ another one, Pro = one, env = <NP, m=}yes\text{, o-m=}yes, O = \text{O-M})}{p(R = \text{ another one, Pro = one, env = <NP, m=}yes\text{, o-m=}yes, O = \text{O-M})}$$

$$= \frac{p(A = \text{red bottle}, i = yes, \text{det} = no, \text{mod} = yes, C = N', R = \text{ another one, Pro = one, env = <NP, m=}yes\text{, o-m=}yes, O = \text{O-M})}{\sum_{\text{A,i,det,mod,C}} p(R = \text{ another one, Pro = one, env = <NP, m=}yes\text{, o-m=}yes, O = \text{O-M})}$$

Pearl & Mis submitted

Context-specific representation: $p_{rep|beh}$

$$p_{rep|beh} = \frac{rep_{1f}}{rep_{1f} + rep_{2f} + rep_{3f}}$$

The outcome where the look to the red bottle is because the learner has the target representation (A="red bottle") and looks at the familiar object.

b = baseline preference for looking at familiar bottle = 0.459
a = adjusted preference for looking at familiar bottle when antecedent is "red bottle" = 0.587

$$rep_{1f} = p_{N'} * \frac{m}{m+n} * p_{incl} * a$$

Category = N', antecedent = "red bottle", adjusted familiarity preference

Pearl & Mis submitted

Context-specific representation: $p_{rep|beh}$

$$p_{rep|beh} = \frac{rep_{1f}}{rep_{1f} + rep_{2f} + rep_{3f}}$$

The outcome where the look to the red bottle is because the learner has the target representation (A="red bottle") and looks at the familiar bottle.

+ Additional outcomes where learner looks at familiar bottle.

b = baseline preference for looking at familiar bottle = 0.459
a = adjusted preference for looking at familiar bottle when antecedent is "red bottle" = 0.587

$$rep_{2f} = p_{N'} * \frac{n}{m+n} * (1 - p_{incl}) * b$$

Category = N', antecedent = "bottle", baseline familiarity preference

$$rep_{3f} = (1 - p_{N'}) * (1 - p_{incl}) * b$$

Category = N^0, antecedent = "bottle", baseline familiarity preference

Pearl & Mis submitted

Learner results: Strategy comparison

Averages over 1000 simulations, s=10, standard deviations in parentheses.

Note: Target p_{beh} = 0.587, all other target p = 1.000

$p_{N'}$
p_{incl}
p_{beh}
$p_{rep\|beh}$

Pearl & Mis submitted

Learner results: Strategy comparison

Averages over 1000 simulations, *s*=10, standard deviations in parentheses.

Note: Target p_{beh} = 0.587, all other target p = 1.000

	DirUnamb
$p_{N'}$	0.500 (<0.01)
p_{incl}	0.500 (<0.01)
p_{beh}	
$p_{rep\|beh}$	

How does a learner who only looks at direct unambiguous evidence fare?

Since the input data include no DirUnamb data, and those are the only data the DirUnamb learner learns from, it learns nothing.

Pearl & Mis submitted

Learner results: Strategy comparison

Averages over 1000 simulations, s=10, standard deviations in parentheses.

Note: Target p_{beh} = 0.587, all other target p = 1.000

	DirUnamb
$p_{N'}$	0.500 (<0.01)
p_{incl}	0.500 (<0.01)
p_{beh}	0.475 (<0.01)
$p_{rep\|beh}$	0.158 (<0.01)

It is at chance for having the target syntactic and referential knowledge necessary to choose the correct antecedent.

It does not generate the observed toddler looking preference, and it is unlikely to have the target representation if it looks at the familiar bottle.

Pearl & Mis submitted

Learner results: Strategy comparison

Averages over 1000 simulations, *s*=10, standard deviations in parentheses.

Note: Target p_{beh} = 0.587, all other target p = 1.000

	DirUnamb
$p_{N'}$	0.500 (<0.01)
p_{incl}	0.500 (<0.01)
p_{beh}	0.475 (<0.01)
$p_{rep\|beh}$	0.158 (<0.01)

Implication:
The learner needs something additional to solve this learning problem.

Pearl & Mis submitted

Learner results: Strategy comparison

Averages over 1000 simulations, *s*=10, standard deviations in parentheses.

Note: Target p_{beh} = 0.587, all other target p = 1.000

	DirUnamb	DirUnamb + N'
$p_{N'}$	0.500 (<0.01)	1.000
p_{incl}	0.500 (<0.01)	
p_{beh}	0.475 (<0.01)	
$p_{rep\|beh}$	0.158 (<0.01)	

What if the learner also knows that *one* is N'? (Baker 1978)

Pearl & Mis submitted

Learner results: Strategy comparison

Averages over 1000 simulations, *s*=10, standard deviations in parentheses.

Note: Target p_{beh} = 0.587, all other target p = 1.000

	DirUnamb	DirUnamb + N'
$p_{N'}$	0.500 (<0.01)	1.000
p_{incl}	0.500 (<0.01)	0.500 (<0.01)
p_{beh}	0.475 (<0.01)	0.492 (<0.01)
$p_{rep\|beh}$	0.158 (<0.01)	0.306 (<0.01)

The DirUnamb + N' learner still has no data to learn the correct referential knowledge.

This lack of referential knowledge causes it not to generate the observed toddler looking preference in context, and even if it happens to look at the familiar bottle, to be unlikely to have the target representation when doing so.

Pearl & Mis submitted

Learner results: Strategy comparison

Averages over 1000 simulations, *s*=10, standard deviations in parentheses.

Note: Target p_{beh} = 0.587, all other target p = 1.000

	DirUnamb	DirUnamb + N’
$p_{N'}$	0.500 (<0.01)	1.000
p_{incl}	0.500 (<0.01)	0.500 (<0.01)
p_{beh}	0.475 (<0.01)	0.492 (<0.01)
$p_{rep\|beh}$	0.158 (<0.01)	0.306 (<0.01)

Implication: Knowing *one* is category N’ isn’t sufficient to generate target behavior if only DirUnamb data are informative.

This learning strategy is insufficient to explain the observed behavior.

Pearl & Mis submitted

Learner results: Strategy comparison

Averages over 1000 simulations, s=10, standard deviations in parentheses.

Note: Target p_{beh} = 0.587, all other target p = 1.000

	DirUnamb	DirUnamb + N'	DirFiltered
$p_{N'}$	0.500 (<0.01)	1.000	0.991 (<0.01)
p_{incl}	0.500 (<0.01)	0.500 (<0.01)	0.963 (<0.01)
p_{beh}	0.475 (<0.01)	0.492 (<0.01)	
$p_{rep\|beh}$	0.158 (<0.01)	0.306 (<0.01)	

Other learning strategies: DirFiltered learner (Regier & Gahl 2004, Pearl & Lidz 2009)

This learner believes *one* is N' when it is smaller than NP and a mentioned property should be included in the antecedent, which is similar to previous findings for this learner.

Pearl & Mis submitted

Learner results: Strategy comparison

Averages over 1000 simulations, *s*=10, standard deviations in parentheses.

Note: Target p_{beh} = 0.587, all other target p = 1.000

	DirUnamb	DirUnamb + N'	DirFiltered
$p_{N'}$	0.500 (<0.01)	1.000	0.991 (<0.01)
p_{incl}	0.500 (<0.01)	0.500 (<0.01)	0.963 (<0.01)
p_{beh}	0.475 (<0.01)	0.492 (<0.01)	0.574 (<0.01)
$p_{rep\|beh}$	0.158 (<0.01)	0.306 (<0.01)	0.918 (<0.01)

Other learning strategies: DirFiltered learner (Regier & Gahl 2004, Pearl & Lidz 2009)

In addition, it is close to generating the observed toddler looking preference, and is likely to have the target representation when looking at the familiar bottle. This new finding suggests this is a pretty successful learning strategy for matching the available behavioral data.

Pearl & Mis submitted

Learner results: Strategy comparison

Averages over 1000 simulations, *s*=10, standard deviations in parentheses.

Note: Target p_{beh} = 0.587, all other target p = 1.000

	DirUnamb	DirUnamb + N'	DirFiltered	DirEO
$p_{N'}$	0.500 (<0.01)	1.000	0.991 (<0.01)	0.246 (0.03)
p_{incl}	0.500 (<0.01)	0.500 (<0.01)	0.963 (<0.01)	0.379 (0.05)
p_{beh}	0.475 (<0.01)	0.492 (<0.01)	0.574 (<0.01)	
$p_{rep\|beh}$	0.158 (<0.01)	0.306 (<0.01)	0.918 (<0.01)	

Other learning strategies: DirEO learner (Pearl & Lidz 2009)

The learner prefers *one* to be N^0 when it is smaller than NP, and does not believe the mentioned property should be included in the antecedent. Neither of these is the target knowledge.

Pearl & Mis submitted

Learner results: Strategy comparison

Averages over 1000 simulations, *s*=10, standard deviations in parentheses.

Note: Target p_{beh} = 0.587, all other target p = 1.000

	DirUnamb	DirUnamb + N'	DirFiltered	DirEO
$p_{N'}$	0.500 (<0.01)	1.000	0.991 (<0.01)	0.246 (0.03)
p_{incl}	0.500 (<0.01)	0.500 (<0.01)	0.963 (<0.01)	0.379 (0.05)
p_{beh}	0.475 (<0.01)	0.492 (<0.01)	0.574 (<0.01)	0.464 (<0.01)
$p_{rep\|beh}$	0.158 (<0.01)	0.306 (<0.01)	0.918 (<0.01)	0.050 (0.01)

Other learning strategies: DirEO learner (Pearl & Lidz 2009)

This causes the learner not to generate the observed toddler looking preference, and not to have the target representation if it looks at the familiar bottle.

Implication: This is not a successful learning strategy for explaining toddler behavior.

Pearl & Mis submitted

Learner results: Strategy comparison

Averages over 1000 simulations, *s*=10, standard deviations in parentheses.

Note: Target p_{beh} = 0.587, all other target p = 1.000

	DirUnamb	DirUnamb + N'	DirFiltered	DirEO	IndirPro
$p_{N'}$	0.500 (<0.01)	1.000	0.991 (<0.01)	0.246 (0.03)	0.368 (0.04)
p_{incl}	0.500 (<0.01)	0.500 (<0.01)	0.963 (<0.01)	0.379 (0.05)	1.000 (<0.01)
p_{beh}	0.475 (<0.01)	0.492 (<0.01)	0.574 (<0.01)	0.464 (<0.01)	
$p_{rep\|beh}$	0.158 (<0.01)	0.306 (<0.01)	0.918 (<0.01)	0.050 (0.01)	

The IndirPro learner robustly decides the antecedent should include the mentioned property.

However, this learner has a moderate dispreference for believing *one* is N' when it is smaller than NP.

This is therefore not the target representation, w.r.t syntactic category.

Pearl & Mis submitted

Learner results: Strategy comparison

Averages over 1000 simulations, *s*=10, standard deviations in parentheses.

Note: Target p_{beh} = 0.587, all other target p = 1.000

	DirUnamb	DirUnamb + N'	DirFiltered	DirEO	IndirPro
$p_{N'}$	0.500 (<0.01)	1.000	0.991 (<0.01)	0.246 (0.03)	0.368 (0.04)
p_{incl}	0.500 (<0.01)	0.500 (<0.01)	0.963 (<0.01)	0.379 (0.05)	1.000 (<0.01)
p_{beh}	0.475 (<0.01)	0.492 (<0.01)	0.574 (<0.01)	0.464 (<0.01)	0.587 (<0.01)
$p_{rep\|beh}$	0.158 (<0.01)	0.306 (<0.01)	0.918 (<0.01)	0.050 (0.01)	0.998 (<0.01)

However…this learner still generates the observed toddler looking preference perfectly, and has the target representation when looking at the familiar bottle.

Pearl & Mis submitted

Learner results: Strategy comparison

Averages over 1000 simulations, *s*=10, standard deviations in parentheses.

Note: Target p_{beh} = 0.587, all other target p = 1.000

	DirUnamb	DirUnamb + N'	DirFiltered	DirEO	IndirPro
$p_{N'}$	0.500 (<0.01)	1.000	0.991 (<0.01)	0.246 (0.03)	0.368 (0.04)
p_{incl}	0.500 (<0.01)	0.500 (<0.01)	0.963 (<0.01)	0.379 (0.05)	1.000 (<0.01)
p_{beh}	0.475 (<0.01)	0.492 (<0.01)	0.574 (<0.01)	0.464 (<0.01)	0.587 (<0.01)
$p_{rep\|beh}$	0.158 (<0.01)	0.306 (<0.01)	0.918 (<0.01)	0.050 (0.01)	0.998 (<0.01)

Why?

The learner believes very strongly that the mentioned property must be included in the antecedent.

Only one antecedent allows this: $[_{N'}$ red$[_{N'}[_{N0}$ bottle$]]]$

Pearl & Mis submitted

Learner results: Strategy comparison

Averages over 1000 simulations, *s*=10, standard deviations in parentheses.

Note: Target p_{beh} = 0.587, all other target p = 1.000

	DirUnamb	DirUnamb + N'	DirFiltered	DirEO	IndirPro
$p_{N'}$	0.500 (<0.01)	1.000	0.991 (<0.01)	0.246 (0.03)	0.368 (0.04)
p_{incl}	0.500 (<0.01)	0.500 (<0.01)	0.963 (<0.01)	0.379 (0.05)	1.000 (<0.01)
p_{beh}	0.475 (<0.01)	0.492 (<0.01)	0.574 (<0.01)	0.464 (<0.01)	0.587 (<0.01)
$p_{rep\|beh}$	0.158 (<0.01)	0.306 (<0.01)	0.918 (<0.01)	0.050 (0.01)	0.998 (<0.01)

Why?

So, because the antecedent includes the mentioned property, it and the pronoun referring to it (*one*) must be N' *in this context* - even if the learner believes *one* is not N' in general.

Pearl & Mis submitted

Learner results: Strategy comparison

Averages over 1000 simulations, s=10, standard deviations in parentheses.

Note: Target p_{beh} = 0.587, all other target p = 1.000

	DirUnamb	DirUnamb + N'	DirFiltered	DirEO	IndirPro
$p_{N'}$	0.500 (<0.01)	1.000	0.991 (<0.01)	0.246 (0.03)	0.368 (0.04)
p_{incl}	0.500 (<0.01)	0.500 (<0.01)	0.963 (<0.01)	0.379 (0.05)	1.000 (<0.01)
p_{beh}	0.475 (<0.01)	0.492 (<0.01)	0.574 (<0.01)	0.464 (<0.01)	0.587 (<0.01)
$p_{rep\|beh}$	0.158 (<0.01)	0.306 (<0.01)	0.918 (<0.01)	0.050 (0.01)	0.998 (<0.01)

Take away point:

A learner using an indirect positive evidence strategy can generate target behavior without reaching the target knowledge state – instead, this learner has a context-sensitive representation (depending on whether a property was mentioned).

Pearl & Mis submitted

A closer look at the IndirPro outcome

Since p_{beh} matches 18-month-old looking preferences, the IndirPro learner succeeds at generating target behavior in this context.

In fact, it will succeed for any context that has a property mentioned in the potential antecedent (due to p_{incl}).

"Look at that baby penguin! Do you see another one?"

$[_{N'}\ \textit{baby}\ [_{N'}\ [_{N0}\ \textit{penguin}]]]$ ← *one*

BABY PENGUIN

A closer look at the IndirPro outcome

However, it will fail to have the target syntactic representation when no property is mentioned.

“Look at that penguin! Do you see another one?”

[$_{N0}$ *penguin*] ← *one*

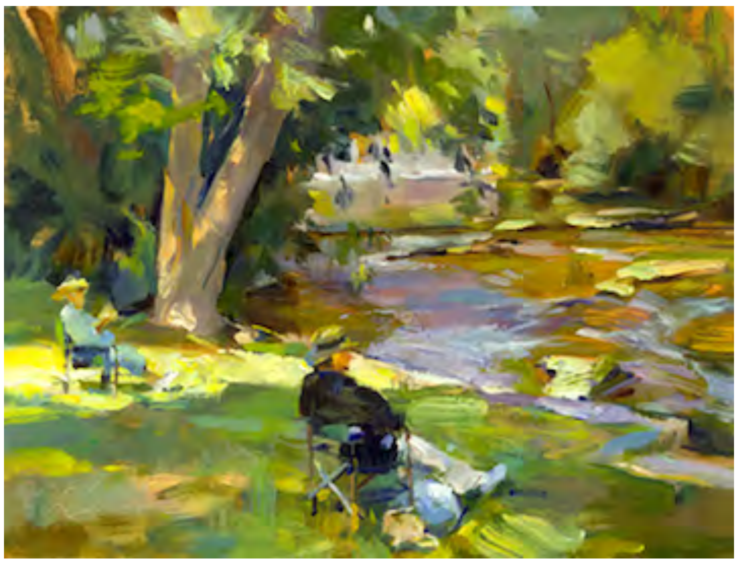
PENGUIN

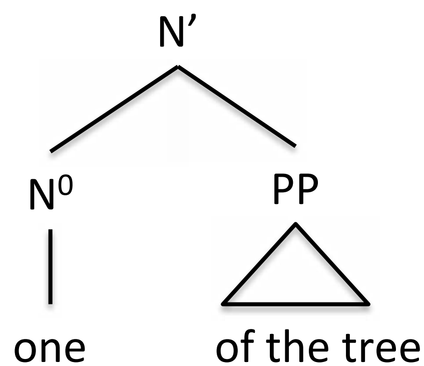
But… this won’t stop it from identifying the correct referent – so communicatively, the learner functions just fine in this context, even with a non-adult syntactic representation.

A closer look at the IndirPro outcome

However, it will allow utterances that adults find ungrammatical, because such utterances use *one* as N^0.

*"I sat by the side of the river when you were sitting by the one of the tree."





So, this is where we would observe a deviation in (use/judgment) behavior from adult behavior. We don't know how 18-month-olds judge these utterances, though.

So what does this mean for learning how to make syntactic generalizations?

The adult generalizations are not necessary to generate the observed 18-month-old behavior. However, *some* syntactic generalizations have been made (maybe the adult ones, but maybe not) and it's important to understand how these could be made.

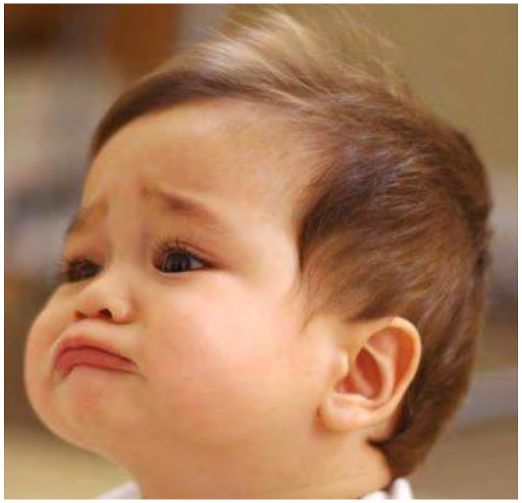
Goal: Learn the appropriate what by the appropriate when using some kind of cognitively plausible how and the available input.

Learning strategy comparison

Unsuccessful strategies for generating target behavior:

✧ DirUnamb + N' (Baker 1978)

- Prior syntactic knowledge is insufficient if only direct positive unambiguous evidence is used.
- Surprising!

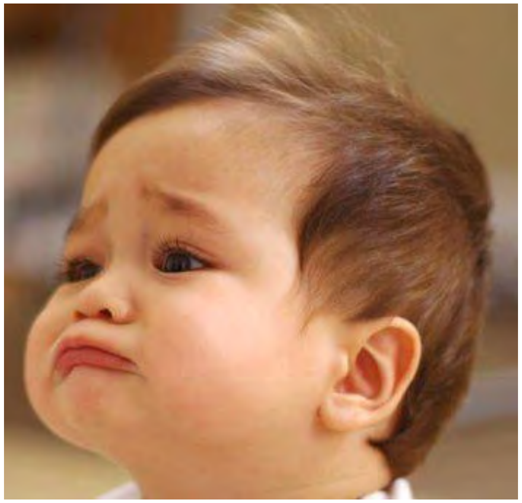


Learning strategy comparison

Unsuccessful strategies for generating target behavior:

✧ DirUnamb + N’ (Baker 1978)

- Prior syntactic knowledge is insufficient if only direct positive unambiguous evidence is used.
- Surprising!



✧ DirEO (Regier & Gahl 2004, Pearl & Lidz 2009)

- Probabilistic inference leverages harmful as well as helpful information from all the direct positive evidence.

Learning strategy comparison

Successful strategies for generating target behavior:

✧ DirFiltered (Regier & Gahl 2004, Pearl & Lidz 2009)

- Probabilistic inference works if certain ambiguous data in the direct positive evidence are filtered out.
- Adult syntactic generalizations are made.

Learning strategy comparison

Successful strategies for generating target behavior:

✧ DirFiltered (Regier & Gahl 2004, Pearl & Lidz 2009)

- Probabilistic inference works if certain ambiguous data in the direct positive evidence are filtered out.
- Adult syntactic generalizations are made.

✧ IndirPro (Pearl & Mis 2011, 2013, submitted)

- Probabilistic inference works if indirect positive evidence coming from other pronoun data is used along with the available direct positive evidence.
- Some non-adult syntactic generalizations are made.

Learning strategy comparison

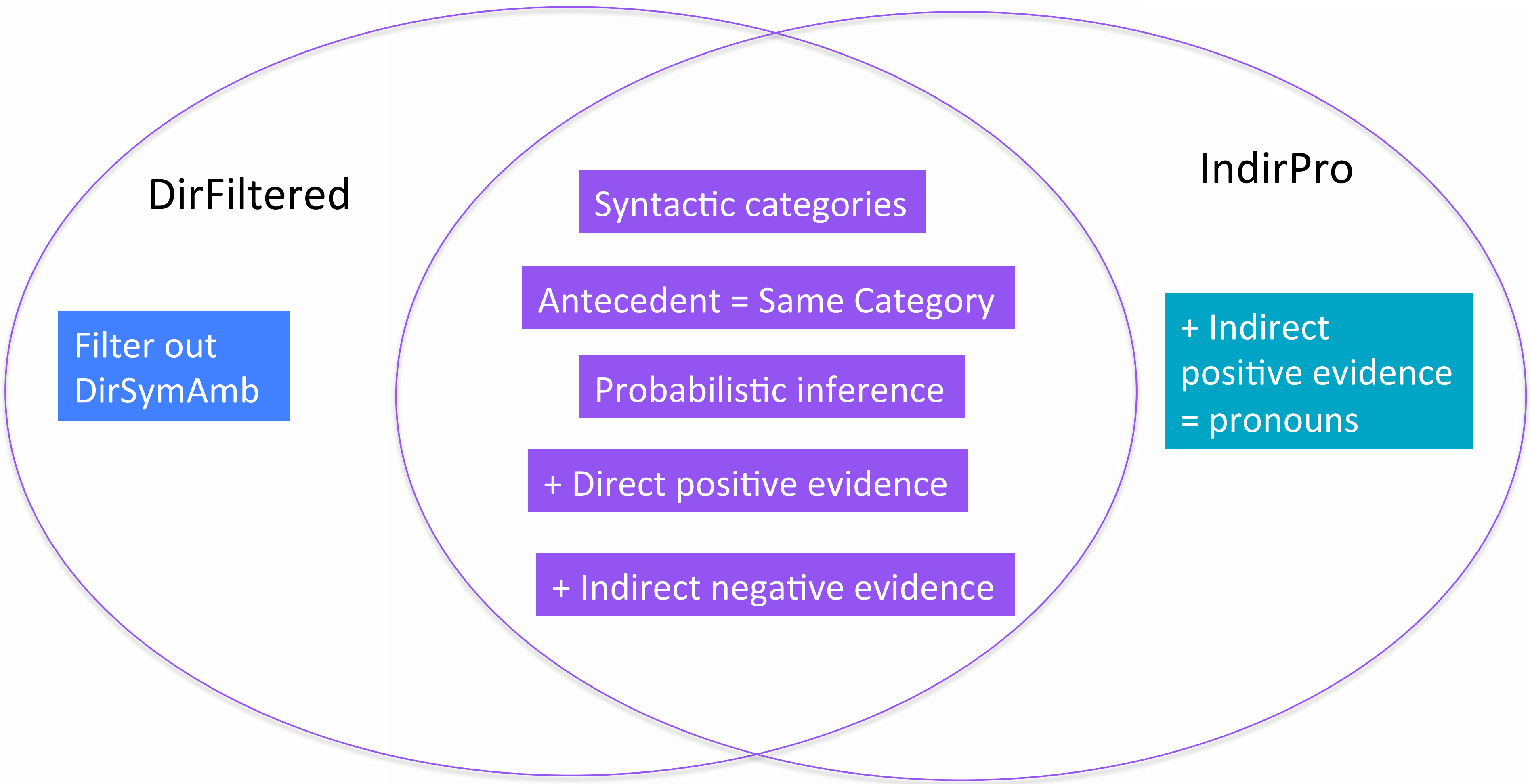
Successful strategies for generating target behavior:

✧ DirFiltered (Regier & Gahl 2004, Pearl & Lidz 2009)

✧ IndirPro (Pearl & Mis 2011, 2013, submitted)

❖ Note: IndirPro more robust than DirFiltered – does not depend on value of *s* being high enough (i.e., the learner finding it highly suspicious that the referent happens to have the mentioned property).

Successful strategy components



Successful strategy components

Both DirFiltered and IndirPro (in fact, all strategies):

Syntactic categories

❖ Knowledge: Syntactic categories exist, in particular N^0, N', and NP.

May be derivable from distributional clustering techniques (e.g., *frequent frames*: Mintz 2003)

or

May require innate, domain-specific knowledge about the kinds of categories that exist in human languages (e.g., existence of N' vs. N^0 is part of Universal Grammar)

Successful strategy components

Both DirFiltered and IndirPro (in fact, all strategies): Antecedent = Same Category

- ❖ Knowledge: Anaphoric elements take linguistic antecedents of the same category.

May be derivable from statistical learning techniques (e.g., probabilistic inference over referential expressions and their linguistic antecedents in unambiguous situations)

or

May require innate, domain-specific knowledge about the relationships that exist between elements in human languages (e.g., Universal Grammar)

Successful strategy components

Both DirFiltered and IndirPro:

Probabilistic inference

❖ Ability: Probabilistic inference

Likely an innate, domain-general ability.

Successful strategy components

Both DirFiltered and IndirPro:

+ Direct positive evidence

- Knowledge: Learn from direct positive evidence available.

Likely innate, domain-general knowledge.

Successful strategy components

Both DirFiltered and IndirPro:

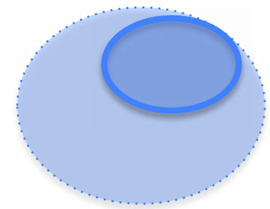
+ Indirect negative evidence

❖ Knowledge: Learn from indirect negative evidence available.

Likely innate, domain-general knowledge, related to probabilistic inference.

Successful strategy components

DirFiltered:



Filter out DirSymAmb

❖ Knowledge: Filter out the DirSynAmb data from the data intake.

Ignore these:

"Look – a bottle!  Do you see another one?" 

$[_{N0}$ *bottle*$]$

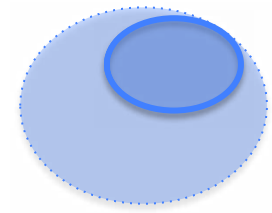
$[_{N'}$ $[_{N0}$ *bottle*$]]$

BOTTLE

Pearl & Lidz (2009) suggest that this filter can be derived from a preference for learning only when there is uncertainty about the referent, as opposed to when there is just uncertainty about the syntactic representation.

Successful strategy components

DirFiltered:



Filter out DirSymAmb

❖ Knowledge: Filter out the DirSynAmb data from the data intake.

Ignore these:

"Look – a bottle!

Do you see another one?"

[$_{N0}$ *bottle*]

[$_{N'}$ [$_{N0}$ *bottle*]]

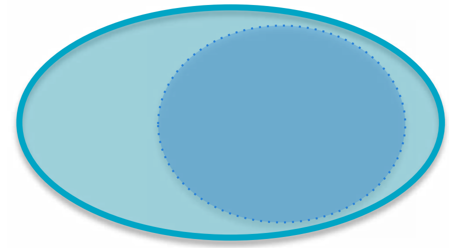
BOTTLE

Open question: Where does this bias for referential over syntactic uncertainty come from?

- Universal Grammar
- Derived from some kind of bias for communicative efficacy (e.g. pay attention if there's ambiguity in understanding, otherwise ignore)

Successful strategy components

IndirPro:



+ Indirect positive evidence = pronouns

❖ Knowledge: Allow in indirect positive evidence from other pronouns.

Include these:
"Look – a blue bottle! Do you want it?"

$[_{NP}\ \textit{a}\ [_{N'}\ \textit{blue}\ [_{N'}\ [_{N0}\ \textit{bottle}]]$

BLUE BOTTLE

This domain-specific knowledge could be specified innately in Universal Grammar.

Referential expressions

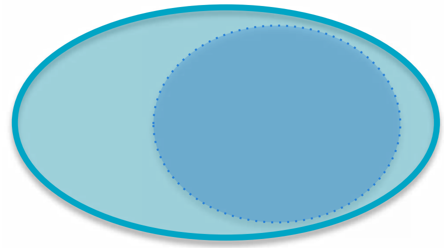
Pronoun_1 Anaphoric *one*

↑

Data_1

Successful strategy components

IndirPro:



+ Indirect positive evidence = pronouns

❖ Knowledge: Allow in indirect positive evidence from other pronouns.

Include these:
"Look – a blue bottle! Do you want it?"

$[_{NP}$ *a* $[_{N'}$ *blue* $[_{N'}$ $[_{N0}$ *bottle*]]

BLUE BOTTLE

But maybe it's derived from observing distributional similarities between anaphoric *one* and other pronouns.

Referential expressions

Pronoun_1 → Anaphoric *one*

$\uparrow$

Data_1

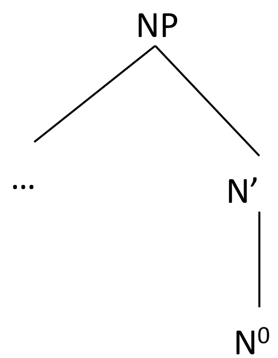
Do you want it?

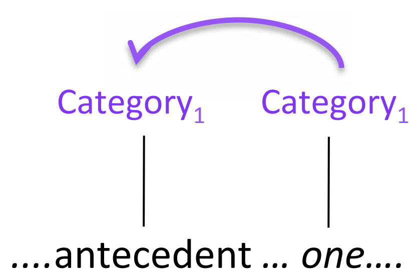
Do you want one?

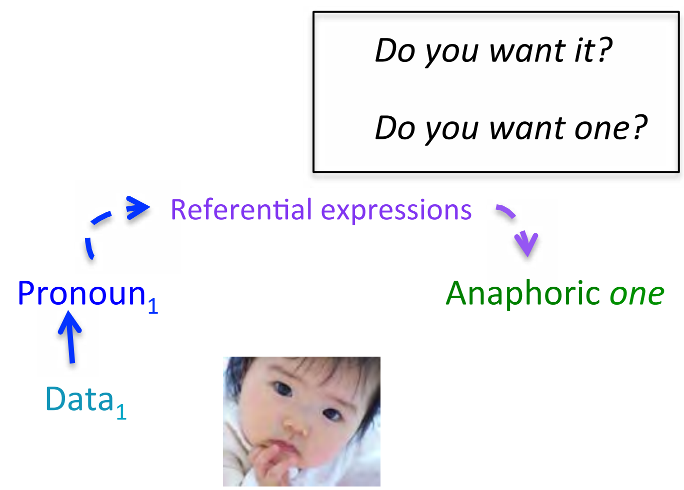
Some open questions

Origin of learning strategy components

For each component that may be derivable from the input, can we create a learner than can actually derive that component from the available linguistic information? And if so, what are the learning components required to do so?







Some open questions

Utility of learning strategy components

How general-purpose are these learning components? Are the components we find useful for making syntactic generalizations about anaphoric *one* useful for making other syntactic generalizations?

Syntactic categories

Probabilistic inference

+ Indirect negative evidence

Some open questions

Utility of learning strategy components

How general-purpose are these learning components? Are the components we find useful for making syntactic generalizations about anaphoric *one* useful for making other syntactic generalizations?

Syntactic categories

Probabilistic inference

+ Indirect negative evidence

What about more generalized forms of those components?

Communicative efficacy
=
Ignore DirSynAmb data

+ Indirect
positive evidence
= pronouns

????

anaphoric *one*

syntactic islands

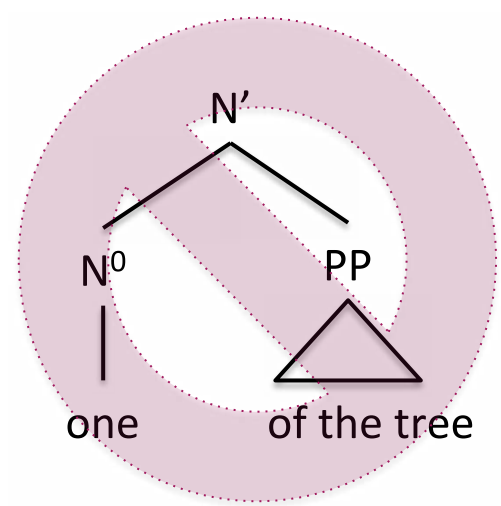
????

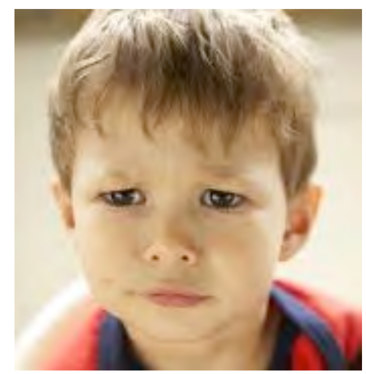
Some open questions

Adult knowledge as target state

Since 18-month-old behavior is consistent with both adult and non-adult syntactic generalizations, how early does the observable behavior occur that is consistent with only adult syntactic generalizations? What knowledge and capabilities are available at that age?

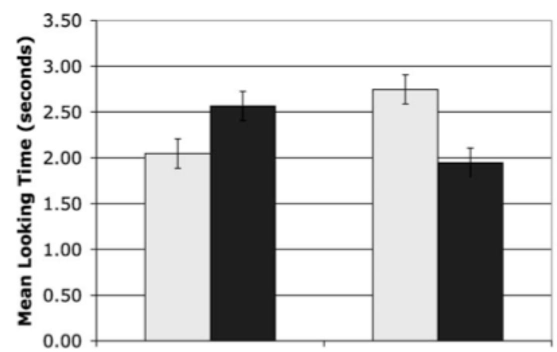
*"I sat by the side of the river when you were sitting by the one of the tree."



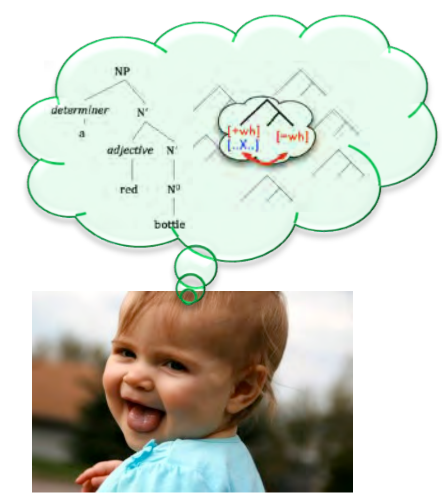


Big picture:
Understanding how children make syntactic generalizations

Target state: What syntactic generalizations are they making?



Empirical data coming from observable behavior is one way to define the goal of learning. This behavior is generated by some underlying syntactic generalizations – maybe not the adult ones (yet), though.

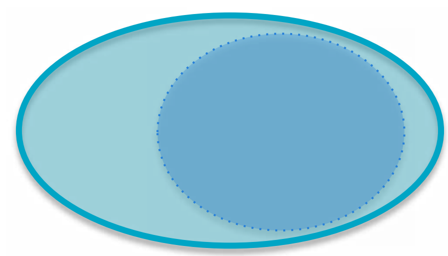


Important: Identifying the generalizations that can produce the observable behavior.

Big picture:
Understanding how children make syntactic generalizations

Indirect positive evidence

If children are probabilistic learners, they may try to leverage any data they perceive as informative. Instead of restricting their input, they may be expanding it beyond the direct evidence in order to make syntactic generalizations.

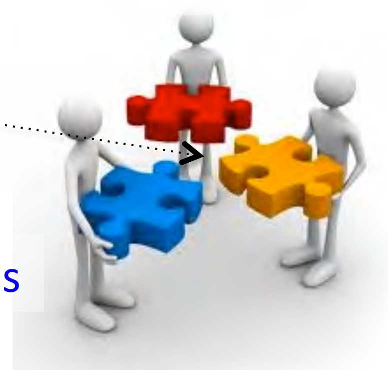


+ Indirect
positive evidence
= pronouns

Big picture:
Understanding how children make syntactic generalizations

Precisely defining the components of any learning problem is necessary for making progress on how children solve that learning problem, which requires the insights from many different methods.

Given a specific initial state, a learner must use the data intake to reach the target state by the end of the learning period.



Thank you!

Benjamin Mis Anousheh Haghighi Jeff Lidz

Jon Sprouse LouAnn Gerken

Max Bane Sue Braunwald Greg Kobele Morgan Sondregger Ming Xiang

Audiences at:

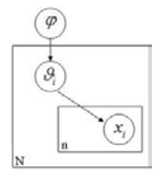
CogSci 2011

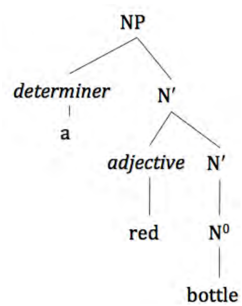
UChicago 2011 workshops on

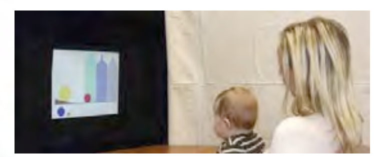
Language, Cognition, and Computation &

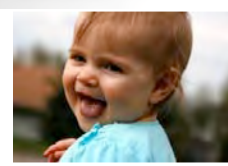
Language, Variation, and Change

NYU Linguistics









Computation of Language Laboratory

UC Irvine

Extra Material

Learner results: Strategy comparison

Averages over 1000 simulations, standard deviations in parentheses.

s = 2, 5, 7, 10, 20, 49   Note: Target p_{beh} = 0.587, all other target p = 1.000

	DirUnamb
$p_{N'}$	0.500 (<0.01)
p_{incl}	0.500 (<0.01)
p_{beh}	0.475 (<0.01)
$p_{rep\|beh}$	0.158 (<0.01)

Since the input data include no DirUnaamb data, and those are the only data the DirUnamb learner learns from, it learns nothing.

It is at chance for having the target syntactic and referential representation.

It will not generate the observed toddler looking preference, and when it does, it unlikely to have the target representation when doing so.

Implication: This learner needs something else if only DirUnamb data are relevant.

Pearl & Mis submitted

Learner results: Strategy comparison

Averages over 1000 simulations, standard deviations in parentheses.

s = 2, 5, 7, 10, 20, 49 Note: Target p_{beh} = 0.587, all other target p = 1.000

	DirUnamb	DirUnamb + N'
$p_{N'}$	0.500 (<0.01)	1.000 (<0.01)
p_{incl}	0.500 (<0.01)	0.500 (<0.01)
p_{beh}	0.475 (<0.01)	0.492 (<0.01)
$p_{rep\|beh}$	0.158 (<0.01)	0.306 (<0.01)

Even if the learner already knows *one* must be category N', there are no data it can use to learn the appropriate referent in this context, which leaves it at chance.

This lack of semantic knowledge causes it not to generate the observed toddler looking preference, and when it does, to be unlikely to have the target representation.

Implication: Knowing *one* is category N' isn't sufficient to generate target behavior if only DirUnamb data are relevant.

Pearl & Mis submitted

Learner results: Strategy comparison

Averages over 1000 simulations, standard deviations in parentheses.

s = 2, 5, 7, 10, 20, 49   Note: Target p_{beh} = 0.587, all other target p = 1.000

	DirUnamb	DirUnamb + N'	IndirPro
$p_{N'}$	0.500	1.000	0.342-0.376
p_{incl}	0.500	0.500	0.998-1.000
p_{beh}	0.475	0.492	0.584-0.587
$p_{rep\|beh}$	0.158	0.306	0.980-1.000

The learner robustly decides the antecedent should include the mentioned property.

However, the learner has a moderate dispreference for believing *one* is N' when it is smaller than NP.

This is therefore not the target representation, w.r.t. syntactic category.

Pearl & Mis submitted

Learner results: Strategy comparison

Averages over 1000 simulations, standard deviations in parentheses.

s = 2, 5, 7, 10, 20, 49 Note: Target p_{beh} = 0.587, all other target p = 1.000

	DirUnamb	DirUnamb + N'	IndirPro
$p_{N'}$	0.500	1.000	0.342-0.376
p_{incl}	0.500	0.500	0.998-1.000
p_{beh}	0.475	0.492	0.584-0.587
$p_{rep\|beh}$	0.158	0.306	0.980-1.000

However…this learner still generates the observed toddler looking preference with high probability, and has the target representation when doing so.

Why? Because the learner believes so strongly that a mentioned property must be included in the antecedent, the only representation that allows this (e.g., $[_{N'}$ red$[_{N'}[_{N0}$ bottle$]]])$ overpowers the other potential representations' probabilities. Thus, the IndirPro learner will conclude the antecedent includes the mentioned property, and so it and the referential pronoun referring to it (one) must be N' *in this context* - even if the learner believes *one* is not N' in general.

Pearl & Mis submitted

Learner results: Strategy comparison

Averages over 1000 simulations, standard deviations in parentheses.

s = 7, 10, 20, 49 Note: Target p_{beh} = 0.587, all other target p = 1.000

	DirUnamb	DirUnamb + N'	DirFiltered	IndirPro
$p_{N'}$	0.500	1.000	0.984-0.995	0.367-0.376
p_{incl}	0.500	0.500	0.906-0.993	0.999-1.000
p_{beh}	0.475	0.492	0.557-0.585	0.586-0.587
$p_{rep\mid beh}$	0.158	0.306	0.807-0.985	0.993-1.000

Other learning strategies: DirFiltered learner (Regier & Gahl 2004, Pearl & Lidz 2009)
Variability, depending on the value of s, which determines how suspicious a coincidence it is that the intended object just happens to have the mentioned property.

When s =7 or above, this learner believes a mentioned property should be included in the antecedent and *one* is N' when it is smaller than NP, which is similar to previous findings by Regier & Gahl 2004 and Pearl & Lidz 2009. In addition, it is likely to generate the observed toddler looking preference, and have the target representation when doing so.

Pearl & Mis submitted

Learner results: Strategy comparison

Averages over 1000 simulations, standard deviations in parentheses.

s = 5 Note: Target p_{beh} = 0.587, all other target p = 1.000

	DirUnamb	DirUnamb + N'	DirFiltered	IndirPro
$p_{N'}$	0.500	1.000	0.942 (<0.01)	0.342 (0.03)
p_{incl}	0.500	0.500	0.683 (<0.01)	0.998 (<0.01)
p_{beh}	0.475	0.492	0.511 (<0.01)	0.584 (<0.01)
$p_{rep\|beh}$	0.158	0.306	0.002 (<0.01)	0.980 (<0.01)

Other learning strategies: DirFiltered learner (Regier & Gahl 2004, Pearl & Lidz 2009)
Variability, depending on the value of s, which determines how suspicious a coincidence it is that the intended object just happens to have the mentioned property.

However, when s=5, the learner is less sure the mentioned property should be included in the antecedent, which causes the learner to be less likely to generate the observed toddler looking preference, and unlikely to have the target representation when doing so.

Pearl & Mis submitted

Learner results: Strategy comparison

Averages over 1000 simulations, standard deviations in parentheses.

s = 2 Note: Target p_{beh} = 0.587, all other target p = 1.000

	DirUnamb	DirUnamb + N'	DirFiltered	IndirPro
$p_{N'}$	0.500	1.000	0.340 (<0.01)	0.342 (0.03)
p_{incl}	0.500	0.500	0.020 (<0.01)	0.998 (<0.01)
p_{beh}	0.475	0.492	0.459 (<0.01)	0.584 (<0.01)
$p_{rep\|beh}$	0.158	0.306	0.000 (<0.01)	0.980 (<0.01)

Other learning strategies: DirFiltered learner (Regier & Gahl 2004, Pearl & Lidz 2009)
Variability, depending on the value of s, which determines how suspicious a coincidence it is that the intended object just happens to have the mentioned property.

When s=2, the learner is sure the mentioned property should *not* be included in the antecedent, and prefer *one* to be N^0 when it is smaller than NP. This causes the learner to not generate the observed toddler looking preference, and not to have the target representation when generating that behavior.

Pearl & Mis submitted

Learner results: Strategy comparison

Averages over 1000 simulations, standard deviations in parentheses.

s = 2, 5 Note: Target p_{beh} = 0.587, all other target p = 1.000

	DirUnamb	DirUnamb + N'	DirFiltered	IndirPro
$p_{N'}$	0.500	1.000	0.340, 0.942	0.342, 0.362
p_{incl}	0.500	0.500	0.020, 0.683	0.998, 0.999
p_{beh}	0.475	0.492	0.459, 0.511	0.584, 0.586
$p_{rep\|beh}$	0.158	0.306	0.000, 0.002	0.980, 0.992

What's going on?

If the suspicious coincidence isn't strong enough, DirRefSynAmb data don't help the learner increase p_{incl} – in fact, they can cause p_{incl} to drop. Because both p_{incl} and $p_{N'}$ are used to calculate ϕ_{incl} and $\phi_{N'}$, a very low p_{incl} can eventually drag $p_{N'}$ down.

Ex: s=2

If the first 20 data points are DirRefSynAmb data points, p_{incl} = 0.12 and $p_{N'}$ = 0.48.

Pearl & Mis submitted

Learner results: Strategy comparison

Averages over 1000 simulations, standard deviations in parentheses.

s = 2, 5, 7, 10 Note: Target p_{beh} = 0.587, all other target p = 1.000

	DirUnamb	DirUnamb + N'	DirFiltered	DirEO	IndirPro
$p_{N'}$	0.500	1.000	0.340-0.991	0.136-0.246	0.367-0.368
p_{incl}	0.500	0.500	0.020-0.963	0.010-0.379	0.999-1.000
p_{beh}	0.475	0.492	0.459-0.574	0.459-0.464	0.586-0.587
$p_{rep\|beh}$	0.158	0.306	0.002-0.918	0.000-0.500	0.993-0.998

Other learning strategies: DirEO learner (Pearl & Lidz 2009)
Variability, depending on the value of s, which determines how suspicious a coincidence it is that the intended object just happens to have the mentioned property.

When s is less than 10, the learner does not believe the mentioned property should be included in the antecedent, and prefers *one* to be N^0 when it is smaller than NP.
This causes the learner to not generate the observed toddler looking preference, and not have the target representation when generating that behavior.

Pearl & Mis submitted

Learner results: Strategy comparison

Averages over 1000 simulations, standard deviations in parentheses.

s = 20, 49 Note: Target p_{beh} = 0.587, all other target p = 1.000

	DirUnamb	DirUnamb + N'	DirFiltered	DirEO	IndirPro
$p_{N'}$	0.500	1.000	0.994, 0.995	0.344, 0.366	0.373, 0.376
p_{incl}	0.500	0.500	0.987, 0.993	0.931, 0.987	1.000
p_{beh}	0.475	0.492	0.582, 0.585	0.532, 0.573	0.587
$p_{rep\|beh}$	0.158	0.306	0.971, 0.985	0.626, 0.912	1.000

Other learning strategies: DirEO learner (Pearl & Lidz 2009)
Variability, depending on the value of s, which determines how suspicious a coincidence it is that the intended object just happens to have the mentioned property.

However, when s is 20 or 49, the learner strongly believes the mentioned property should be included in the antecedent, though it still prefers *one* to be N^0 when it is smaller than NP. This causes the learner to be more likely to generate the observed toddler looking preference, and more likely to have the target representation when generating that behavior.

Pearl & Mis submitted

Learner results: Strategy comparison

Averages over 1000 simulations, standard deviations in parentheses.

s = 20, 49 Note: Target p_{beh} = 0.587, all other target p = 1.000

	DirUnamb	DirUnamb + N'	DirFiltered	DirEO	IndirPro
$p_{N'}$	0.500	1.000	0.994, 0.995	0.344, 0.366	0.373, 0.376
p_{incl}	0.500	0.500	0.987, 0.993	0.931, 0.987	1.000
p_{beh}	0.475	0.492	0.582, 0.585	0.532, 0.573	0.587
$p_{rep\|beh}$	0.158	0.306	0.971, 0.985	0.626, 0.912	1.000

Other learning strategies: DirEO learner (Pearl & Lidz 2009)
Variability, depending on the value of s, which determines how suspicious a coincidence it is that the intended object just happens to have the mentioned property.

This is more like the IndirPro learner results.

Pearl & Mis submitted

Learner results: Strategy comparison

Averages over 1000 simulations, standard deviations in parentheses.

s = 20, 49 Note: Target p_{beh} = 0.587, all other target p = 1.000

	DirUnamb	DirUnamb + N'	DirFiltered	DirEO	IndirPro
$p_{N'}$	0.500	1.000	0.994, 0.995	0.344, 0.366	0.373, 0.376
p_{incl}	0.500	0.500	0.987, 0.993	0.931, 0.987	1.000
p_{beh}	0.475	0.492	0.582, 0.585	0.532, 0.573	0.587
$p_{rep\|beh}$	0.158	0.306	0.971, 0.985	0.626, 0.912	1.000

What's going on?

The flip side of what we saw with the DirFiltered learner. If the suspicious coincidence is very strong, DirRefSynAmb data help the learner increase p_{incl} (and $p_{N'}$) – in fact, they become almost as influential as DirUnamb data. Because both p_{incl} and $p_{N'}$ are used to calculate ϕ_{incl} and $\phi_{N'}$, a very high p_{incl} can bolster $p_{N'}$, and mostly overpower the effect of the troublesome DirSynAmb data.

Pearl & Mis submitted

Learner results: Strategy comparison

Averages over 1000 simulations, standard deviations in parentheses.

s = 2, 5, 7, 10, 20, 49 Note: Target p_{beh} = 0.587, all other target p = 1.000

	DirUnamb	DirUnamb + N'	DirFiltered	DirEO	IndirPro
$p_{N'}$	0.500	1.000	0.340-0.995	0.136-0.366	0.342-0.376
p_{incl}	0.500	0.500	0.020-0.993	0.010-0.987	0.998-1.000
p_{beh}	0.475	0.492	0.459-0.585	0.459-0.573	0.584-0.587
$p_{rep\|beh}$	0.158	0.306	0.002-0.985	0.000-0.912	0.980-1.000

Why isn't the IndirPro learner as succeptible to changing s values?

IndirUnamb data only ever increase p_{incl}, no matter what the value of s. So, because there are so many of them, they can overwhelm the effect of DirRefSynAmb data on p_{incl} (whether s is low or high). This helps keep $p_{N'}$ from plummeting, though it still drops due to the troublesome DirSynAmb data in the learner's intake.

Pearl & Mis submitted

Learner results: Strategy comparison

Averages over 1000 simulations, standard deviations in parentheses.

s = 2, 5, 7, 10, 20, 49 Note: Target p_{beh} = 0.587, all other target p = 1.000

	DirUnamb	DirUnamb + N'	DirFiltered	DirEO	IndirPro
$p_{N'}$	0.500	1.000	0.340-0.995	0.136-0.366	0.342-0.376
p_{incl}	0.500	0.500	0.020-0.993	0.010-0.987	0.998-1.000
p_{beh}	0.475	0.492	0.459-0.585	0.459-0.573	0.584-0.587
$p_{rep\|beh}$	0.158	0.306	0.002-0.985	0.000-0.912	0.980-1.000

Take away points:.

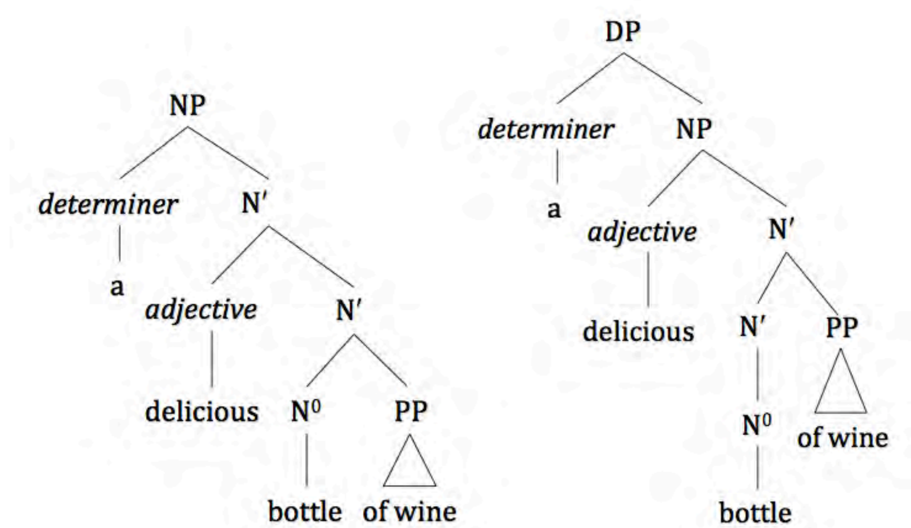
An indirect positive evidence learning strategy has a beneficial impact on learning anaphoric *one* – it makes the learner's behavior robust, no matter how suspicious a coincidence the DirRefSynAmb data are (or aren't).

A learner using an indirect positive evidence strategy can generate target behavior without reaching the target knowledge state – instead, this learner has a context-sensitive representation (depending on whether a property was mentioned).

Pearl & Mis submitted

An alternate theoretical representation

N^0, N', and NP vs. N^0, N', NP, and DP



An alternate theoretical representation

Initial state

- **Knowledge: Syntactic categories exist, in particular N^0, N', NP, and DP.**
- Knowledge: Anaphoric elements like *one* take linguistic antecedents of the same category.

Target state

- Knowledge: In utterances like "Look - a red bottle! Look - another one!", ***one* is category NP** and so its antecedent includes the modifier ("red").
- Behavior: In the LWF experiment, the learner should look at the familiar (red) bottle with a higher probability.

An alternate theoretical representation

What an indirect positive evidence strategy like IndirPro would do

Initial state

- **Knowledge: Syntactic categories exist, in particular N^0, N', NP, and DP.**
- Knowledge: Anaphoric elements like *one* take linguistic antecedents of the same category.

+ Direct positive, indirect negative, and indirect positive data are informative.
+ Indirect positive evidence = other referential pronoun data
+ Use probabilistic inference

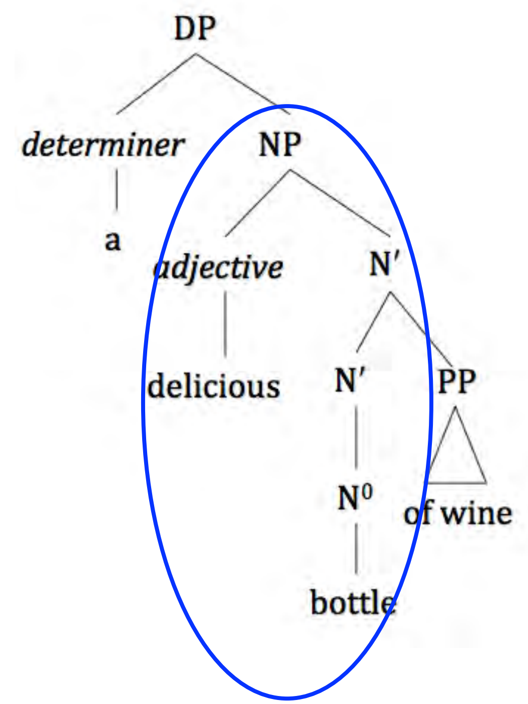
An alternate theoretical representation

What an indirect positive evidence strategy like IndirPro would do

(1) DirUnamb data still indicate antecedent that includes modifier – it's just that the category label is NP (rather than N').

p_{incl} and p_{NP} both increase.

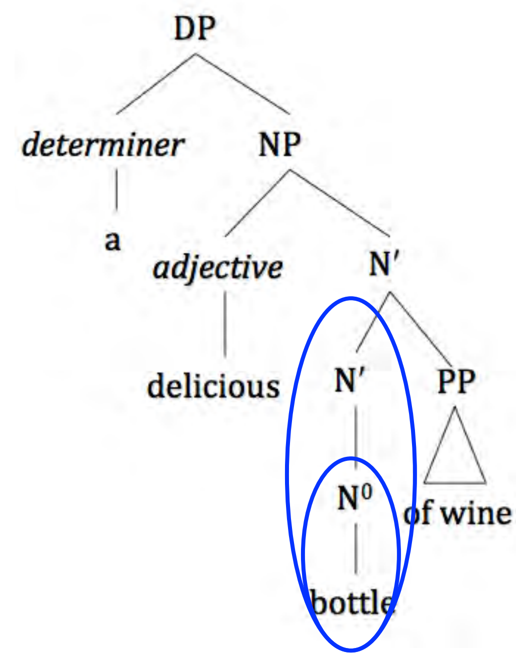
DirUnamb data still cause p_{incl} and the category that includes the modifier (NP) to increase.



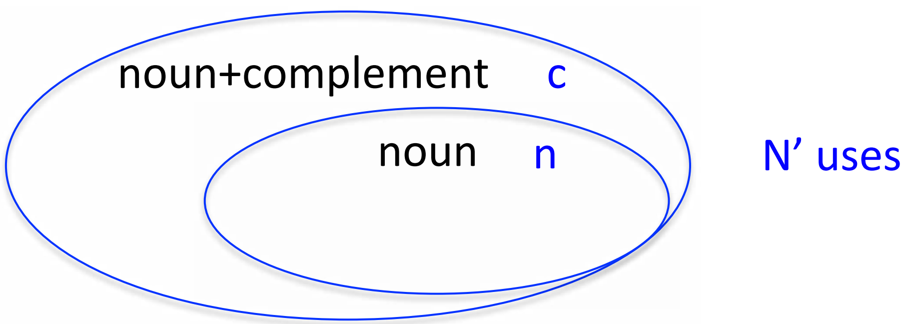
An alternate theoretical representation

What an indirect positive evidence strategy like IndirPro would do

(2) DirSynAmb data still ambiguous between two categories (N^0 and N'), and probabilistic inference causes learner to prefer the hypotheses that includes fewer strings, which is still the N^0 category. (N' includes noun+ complement strings)



DirSynAmb data still cause $p_{N'}$ to drop, though perhaps not as fast, depending on frequency of complements in the learner's input.

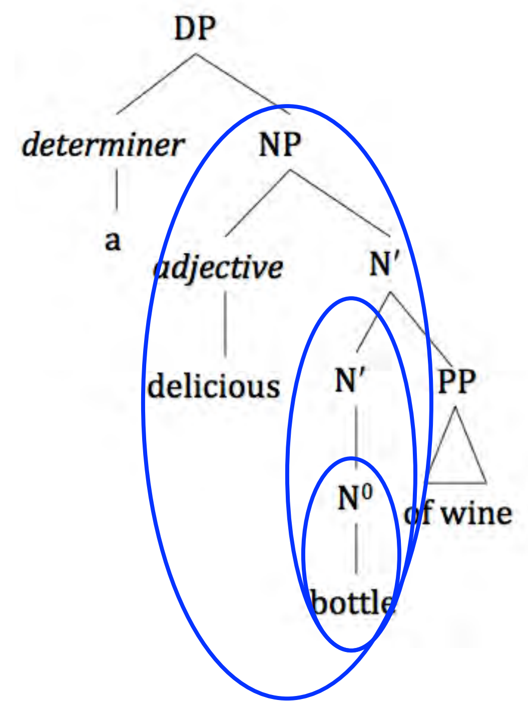


An alternate theoretical representation

What an indirect positive evidence strategy like IndirPro would do

(3) DirRefSynAmb data still ambiguous between three antecedents. When *s* is high enough (>5), the suspicious coincidence still causes the learner to increase p_{incl}.

DirRefSyndata still cause p_{incl} to increase when the suspicious coincidence is strong enough.



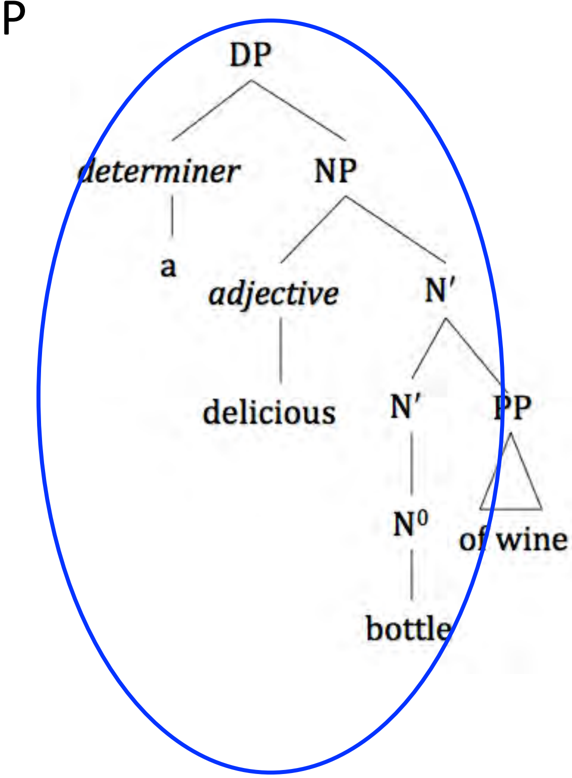
An alternate theoretical representation

What an indirect positive evidence strategy like IndirPro would do

(4) IndirUnamb data still indicate that antecedent includes modifier – it's just that the category label is DP (rather than NP).

p_{incl} still increases.

IndirUnamb data still cause p_{incl} to increase.



An alternate theoretical representation

What an indirect positive evidence strategy like IndirPro would do

Given that the updates from the different data types are effectively the same, the overall outcome should be similar: p_{incl} should be high while p_{NP} should be low.
(Note: $p_{N'}$ should also be very low, since no data cause it to increase.)

Non-target context-dependent representation.
p_{incl} = high, p_{NP} = low

LWF experiment: target behavior (and target representation when displaying that behavior) because of p_{incl}.