Looking Beyond: What Indirect Evidence Can Tell Us About Universal Grammar

Lisa S. Pearl
Workshop on Language, Cognition, and Computation & Workshop on Language, Variation, and Change
University of Chicago
March 4, 2011

An induction problem by any other name...

One of the most controversial claims in linguistics is that children face an induction problem:

"Logical Problem of Language Acquisition" (Baker 1981, Hornstein & Lightfoot 1981)
"Plato's Problem" (Chomsky 1988, Dresher 2003)

Basic claim:
The data encountered are compatible with multiple hypotheses.

The induction problem

Extended claim:
Given this, the data are insufficient for identifying the correct hypothesis as quickly as children do (Logie & Yang 2002) – or at all.

Big question: How do children do it, then?

One answer: Children come prepared

- Children are not unbiased learners.
- But if children come equipped with helpful learning biases, then what is the nature of these necessary biases?
  - Are they innate or derived from the input somehow?
  - Are they domain-specific or domain-general?
  - Are they about what’s being learned or about how to learn?

The Universal Grammar (UG) hypothesis (Chomsky 1965, Chomsky 1975):
These biases are innate and domain-specific.
**Induction problems, UG, and informative data**

Traditional Idea

Traditional assumption:
Only directly related data are informative data. These data are often rare, and that’s why induction problems occur.

**The direct evidence assumption**

If you want to learn linguistic knowledge L, you learn it by observing examples of L in your input (and possibly by also being sensitive to indirect negative evidence about what examples are missing from the input.)

Learning complex yes/no questions

Direct evidence L:
“Is the boy who is in the corner t₁, happy?”

Possible indirect negative evidence:
“Is the boy who t₁ in the corner is happy?”
The direct evidence assumption
If you want to learn linguistic knowledge L, you learn it by observing examples of L in your input (and possibly by also being sensitive to indirect negative evidence about what examples are missing from the input.)

Learning about the distribution of noun phrases (what Case Theory explains):
Direct evidence L: *John seems to be clever.*
*John tries to be clever.*
*It seems John is clever.*

Possible indirect negative evidence:
**"It tries John is clever."**

A broader set of informative data
*Indirect evidence*: other kinds of data that may also be relevant, thereby broadening the set of informative data

Recent computational models have been exploring this:
- Complex yes/no questions (Perfors, Tenenbaum, & Regier 2006, 2011)
- Anaphoric one (Regier & Gahl 2004; Pearl & Lida 2009; Foraker et al. 2009)

Mapping out UG & the acquisition process

Big questions:
- When induction problems exist, what does it take to solve them?
  - What indirect evidence is available? How might a child leverage this evidence?
  - What learning biases can get the job done, given the available data?
    Are they necessarily innate and domain-specific (UG)?
- How can the necessary learning biases inform us about how the acquisition process works?

Three studies of indirect evidence at UC Irvine: *"Testing the Universal Grammar Hypothesis"

Today
Learning the representation of English anaphoric one
"I like the student of linguistics and he likes the one of computer science."

Learning syntactic islands:
"Who did the teacher think [the letter from the teacher] inspired the students?"

Learning about the distribution of noun phrases (what Case Theory explains):
"It tries John is clever."
Road Map

- Adult and child knowledge states for anaphoric one
- The learning problem, given the available data
- Previous proposals for how to solve this problem
- A broader view of informative data
- Representing the information in the data
- An online probabilistic learning framework
- Results & implications

Anaphoric One

Look - a red bottle!

Do you see another one?

Process: First determine the antecedent of one (what string one is referring to). ⇒ "red bottle"

Anaphoric One

Look - a red bottle!

Do you see another one?

red bottle

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
Anaphoric One

Look - a red bottle!

Do you see another one?

Two steps:
1. Identify syntactic antecedent (based on syntactic category of one)
2. Identify semantic referent (based on syntactic antecedent)

Anaphoric One: Syntactic Category

Standard linguistic theory claims that one in these kinds of utterances is a syntactic category smaller than an entire noun phrase, but larger than just a noun (N'). This category is sometimes called N. This category includes strings like "bottle" and "red bottle".

Anaphoric One: Syntactic Category

Importantly, one is not N'. If it was, it could only have strings like "bottle" as its antecedent, and could never have strings like "red bottle" as its antecedent.
**Anaphoric One: Interpretations based on Syntactic Category**

If one was N, we would have a different interpretation of "Look - a red bottle! Do you see another one?"

Because one's antecedent could only be "bottle", we would interpret the second part as "Do you see another bottle?" and the purple bottle would be a fine referent for one.

Since one's antecedent is "red bottle", and "red bottle" cannot be N, one must not be N.

---

**Anaphoric One: Children's Knowledge**

Lidz, Waxman, & Freedman (2002) [LWF] found that 18-month-olds have a preference for the red bottle in the same situation.

"Look - a red bottle! Do you see another one?"

LWF interpretation & conclusion:

- Preference for the red bottle means the preferred syntactic antecedent is "red bottle".

LWF conclude that 18-month-old knowledge = syntactic category of one = N

syntactic antecedent when modifier is present includes modifier (e.g., red) = referent has modifier property

---

**Anaphoric One: The induction problem**

Acquisition: Children must learn the right syntactic category for one, and the right interpretation preference for one in situations with more than one option.

Problem: Most data children encounter are ambiguous.

Syntactically (Sy) ambiguous data:

"Look - a bottle! Oh, look - another one."

one's referent = BOTTLE
one's antecedent = [a [a bottle]] or [one bottle]?
**Anaphoric One: The induction problem**

Acquisition: Children must learn the right syntactic category for one, and the right interpretation preference for one in situations with more than one option.

**Problem:** Most data children encounter are ambiguous.

Semantically and syntactically (Sim-Syn) ambiguous:

“Look - a red bottle! Oh, look - another one.”

one’s referent = RED BOTTLE or BOTTLE?
one’s antecedent = \[\text{red,} \text{bottle}] \text{ or } \[\text{bottle}] \text{ or } \[\text{bottle}]?

---

**Previous proposals for learning about one**


How then?

Children have innate, domain-specific knowledge restricting the hypotheses about one. one cannot be syntactic category N.

What about when there are multiple N’ antecedents?

\[\text{red}\langle \text{bottle} \rangle] \text{ or } \[\text{bottle}]?

(No specific proposal for this.)

---

**Anaphoric One: The induction problem**

Acquisition: Children must learn the right syntactic category for one, and the right interpretation preference for one in situations with more than one option.

**Problem:** Unambiguous data are rare (<0.25%), based on LWF’s analysis.

Unambiguous (Sim-Syn) 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 = \[\text{red,} \text{bottle}]\).

---

**Previous proposals for learning about one**

Regier & Gahl 2004 [R&G]: Sem-Syn ambiguous data can be leveraged, in addition to using unambiguous data.

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

How?

Use innate domain-general statistical learning abilities to track how often one’s referent has the mentioned property (e.g. red). If the referent often has the property (RED BOTTLE), this is a suspicious coincidence unless the antecedent really does include the modifier (“red bottle”) and one’s category is N.

\[\text{red,} \text{bottle} \]
Previous proposals for learning about *one*

Pearl & Liza 2009 [P&L]: Syn ambiguous data must not be leveraged, even if Sem-Syn and unambiguous data are used. “Look - a bottle! Oh, look - another one!”

Why?
These data cause an “equal-opportunity” (EU) probabilistic learner to think *one’s* category is N. [bottle]

How?
P&L propose a domain-specific learning bias to ignore just these ambiguous data, though they speculate how this bias could be derived from an innate domain-general preference for learning when there is local uncertainty.

Foraker et al. 2009 [F&al]: Leverage the syntactic distribution of *one* with innate domain-general statistical learning, by using subtle domain-specific semantic distinctions that indicate syntactic category.

“ball with stripes” “side of the road”
“one with dots” “*one of the river*”
[modifiers] [complements = conceptually evoked by head noun]
[head noun = N] [head noun = N]

How?
Indirect negative evidence (never seeing *one* with a complement, even though other nouns take complements) indicates *one* is not N.

A new proposal: Broadening the data set

Pearl & Mis, submitted [P&M]: Other pronouns in the language can also be used anaphorically: him, her, it ...

Look at the cute penguin. I want to hug him/ her/ it.
[is a cute [an [a penguin]]] → [is him/her/it]

Look! A cute penguin. I want one.
[is a cute [an [a penguin]]] → [is one]

Note: The issue of *one’s* category only occurs when *one* is used in a syntactic environment that indicates it is smaller than an NP (<NP).

Pearl & Mis, submitted [P&M]: Track how often the referent of the anaphoric element (*one, him, her, it, etc.*) has the property mentioned in the potential antecedent, using innate domain-general statistical learning abilities.

Important: This applies, even when the syntactic category is known.

Look at the cute penguin. I want to hug him/ her/ it.
Look! A cute penguin. I want one.

Is the referent cute? Yes! So it’s important that the antecedent include the modifier “cute.”
A new proposal: Broadening the data set

Pearl & Mis, submitted [P&M]: Track how often the referent of the anaphoric element (one, him, her, it, etc.) has the property mentioned in the potential antecedent, using innate domain-general statistical learning abilities.

Important: This applies, even when the syntactic category is known.

Look at the cute penguin. I want to hug him/her/it.

Look! A cute penguin. I want one.

Data points like those above will always include the modifier in the antecedent, since the category of the pronoun is NP and so the antecedent is the entire NP. These data are unambiguous: the referent must have the mentioned property.

Data set comparisons:
Learners using syntactic and semantic information

Unamb <NP
"Look – a red bottle! Hmmm – there doesn’t seem to be another one here, though."

Learners: Baker, R&G, P&L’s EO, P&M

Sem-Syn Amb
"Look – a red bottle! Oh, look – another one!"

Learners: R&G, P&L’s EO, P&M

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

Learners: P&L’s EO, P&M

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

Learners: P&M
Information in the data: Unamb <NP

**previous context =**
ex: "...a red bottle..."

**current usage = pronoun**
ex: "...another one..."

**REFERENTIAL INTENT**
- Property mentioned?: no
- Property important?: no
- Antecedent string includes property?: yes

**SYNTACTIC USAGE**
- Pronoun used: one
- Syntactic category of pronoun: N, N'
- Antecedent string includes modifier?: no
- Syntactic environment: 

**Actual antecedent string**
*red bottle*

**Object referred to**

**Observed**

**Latent**

Information in the data: Sem-Syn Amb

**previous context =**
ex: "...a red bottle..."

**current usage = pronoun**
ex: "...another one..."

**REFERENTIAL INTENT**
- Property mentioned?: no
- Property important?: no
- Antecedent string includes property?: yes

**SYNTACTIC USAGE**
- Pronoun used: one
- Syntactic category of pronoun: N, N'
- Antecedent string includes modifier?: yes
- Syntactic environment: 

**Actual antecedent string**
*red bottle*, *bottle*

**Object referred to**

**Observed**

**Latent**

Information in the data: Syn Amb

**previous context =**
ex: "...a bottle...

**current usage = pronoun**
ex: "...another one..."

**REFERENTIAL INTENT**
- Property mentioned?: no
- Property important?: no
- Antecedent string includes property?: yes

**SYNTACTIC USAGE**
- Pronoun used: one
- Syntactic category of pronoun: N, N'
- Antecedent string includes modifier?: no
- Syntactic environment: 

**Actual antecedent string**
*picture*  

**Object referred to**

**Observed**

**Latent**

Information in the data: Unamb NP

**previous context =**
ex: "...a red bottle..."

**current usage = pronoun**
ex: "...want one..."

**REFERENTIAL INTENT**
- Property mentioned?: yes
- Property important?: yes
- Antecedent string includes property?: yes

**SYNTACTIC USAGE**
- Pronoun used: one
- Syntactic category of pronoun: N, N'
- Antecedent string includes modifier?: yes
- Syntactic environment: 

**Actual antecedent string**
*a red bottle*

**Object referred to**

**Observed**

**Latent**
The online probabilistic framework

Tracking the probability that a property mentioned in the potential antecedent is important: \( p_j \)

- Property mentioned = yes
- Property important?

Tracking the probability that the syntactic category is \( N' \) when it is smaller than NP: \( p_{N'} \)

- Syntactic category of pronoun

Syntactic environment \( = \text{NP} \)

The online probabilistic framework

General form of update equations for \( p_j \) (adapted from Chew 1971):

<table>
<thead>
<tr>
<th>Data seen suggesting ( x ) is true</th>
<th>( \frac{\alpha + \text{data}}{\alpha + \beta \text{total data}} )</th>
<th>A very weak prior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data seen ( = ) total data ( + \frac{\beta}{\alpha} )</td>
<td>Incremented by probability that data point suggests ( x ) is true</td>
<td></td>
</tr>
<tr>
<td>total data ( = ) total data ( + \frac{\beta}{\alpha} )</td>
<td>One informative data point seen</td>
<td></td>
</tr>
</tbody>
</table>

The online probabilistic framework:

Updating \( p_j \)

<table>
<thead>
<tr>
<th>( \phi_i )</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unamb &lt;NP</td>
<td>1</td>
</tr>
<tr>
<td>Unamb NP</td>
<td>1</td>
</tr>
<tr>
<td>Syn Amb</td>
<td>N/A</td>
</tr>
<tr>
<td>Sem-Syn Amb</td>
<td>( \frac{\rho_j}{\rho_j + \rho_1 + \rho_2} )</td>
</tr>
</tbody>
</table>

- \( \rho_j = p_j \frac{m + n}{m + n} \) \( + p_1 \) Category = \( N' \), choose \( N' \) with modifier, property is important
- \( \rho_1 = p_1 \frac{m + n}{m + n} \) \( + (1 - p_1) \) \( \frac{1}{2} \) Category = \( N' \), choose \( N' \) without modifier, property is not important, choose object with property by chance
- \( \rho_2 = (1 - p_0) \frac{(1 - p_0)}{(1 - p_0) + \frac{1}{2}} \) Category = \( N' \), property is not important, choose object with property by chance

The online probabilistic framework:

Updating \( p_{N'} \)

<table>
<thead>
<tr>
<th>( \phi_{N'} )</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unamb &lt;NP</td>
<td>1</td>
</tr>
<tr>
<td>Unamb NP</td>
<td>N/A</td>
</tr>
<tr>
<td>Syn Amb</td>
<td>( \frac{\rho_{N'}}{\rho_{N'} + \rho_1 + \rho_2} )</td>
</tr>
<tr>
<td>Sem-Syn Amb</td>
<td>( \frac{\rho_{N'} + \rho_1}{\rho_{N'} + \rho_1 + \rho_2} )</td>
</tr>
</tbody>
</table>

- \( \rho_{N'} = p_{N'} \frac{m + n}{m + n} \) \( + p_1 \) Category = \( N' \), choose \( N' \) with modifier, property is important
- \( \rho_1 = p_1 \frac{m + n}{m + n} \) \( + (1 - p_1) \) \( \frac{1}{2} \) Category = \( N' \), choose \( N' \) without modifier, property is not important, choose object with property by chance
- \( \rho_2 = (1 - p_0) \frac{(1 - p_0)}{(1 - p_0) + \frac{1}{2}} \) Category = \( N' \), property is not important, choose object with property by chance
### The online probabilistic framework: Updating \( p_N \)

<table>
<thead>
<tr>
<th>( \Phi_N )</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unamb &lt;NP</td>
<td>1</td>
</tr>
<tr>
<td>Unamb NP</td>
<td>N/A</td>
</tr>
<tr>
<td>Syn Amb</td>
<td>( \rho_\psi )</td>
</tr>
<tr>
<td>Sem-Syn Amb</td>
<td>( \rho_\psi )</td>
</tr>
</tbody>
</table>

\[
\rho_\psi = p_N \frac{n + \psi}{m + n + \psi} \quad \text{Category} \quad \text{N', choose N' without modifier}
\]

\[
\rho = 1 - \rho_\psi \quad \text{Category} \quad \text{N''}
\]

### Example updates

Start with \( p_N = p_1 = 0.50 \)

- One Unamb <NP data point: \( p_N = 0.67, p_2 = 0.67 \)
- One Unamb NP data point: \( p_N = 0.50, p_2 = 0.67 \)
- One Sem-Syn Amb data point: \( p_N = 0.56, p_2 = 0.47 \) \( n=1, m=1, r=5 \) [from P&L]
- One Syn Amb data point: \( p_N = 0.48, p_2 = 0.50 \) \( n=1, m=3, r=5 \) [from P&L]

---

### Corpus Analysis & Learner Input

**Brown/Eve corpus (CHILDES; MacWhinney 2000):** starting at 18 months
17,521 utterances of child-directed speech, 2,874 referential pronoun utterances

<table>
<thead>
<tr>
<th></th>
<th>Learner Input based on Brown/Eve corpus distributions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unamb &lt;NP</td>
<td>Baker: 0.00%  R&amp;G: 0  P&amp;L's EO: 0  P&amp;M: 0</td>
</tr>
<tr>
<td>Sem-Syn Amb</td>
<td>Baker: 0.66%  R&amp;G: 242  P&amp;L's EO: 242  P&amp;M: 242</td>
</tr>
<tr>
<td>Syn Amb</td>
<td>Baker: 7.52%  R&amp;G: 0  P&amp;L's EO: 2743  P&amp;M: 2743</td>
</tr>
<tr>
<td>Unamb NP</td>
<td>Baker: 8.42%  R&amp;G: 0  P&amp;L's EO: 0  P&amp;M: 3073</td>
</tr>
<tr>
<td>Uninformative</td>
<td>Baker: 83.4%  R&amp;G: 36500  P&amp;L's EO: 36258  P&amp;M: 33515</td>
</tr>
</tbody>
</table>

Pearl & Lids (2009): Children learn one's representation 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 (36,500 referential pronoun utterances total).
Measures of Success: LWF children's behavior

In addition to directly assessing $p_l$ and $p_m$, we can measure how often a learner would reproduce the behavior in the LWF experiment.

Look – a red bottle!

Do you see another one?

Testing LWF’s assumption about what behavior means

In addition to directly assessing the learner’s behavior, we can assess LWF's assumption that correct behavior indicates the children have the correct representation for one.

Is it possible to get correct behavior in the LWF experiment without having the correct representation for one in general (as measured by $p_l$ and $p_m$)?

Is it possible to get correct behavior in the LWF experiment without having the correct representation for one at the time the behavior is being produced?

\[
\frac{p_l}{p_l + p_m + p_n} \quad \text{given that the learner looks at the red bottle}
\]

\[
\frac{p_m}{p_l + p_m + p_n} \quad \text{given that the learner looks at another bottle}
\]

\[
\frac{p_n}{p_l + p_m + p_n} \quad \text{given that the learner looks at no bottle}
\]

Measures of Success: LWF children's behavior

In addition to directly assessing $p_l$ and $p_m$, we can measure how often a learner would reproduce the behavior in the LWF experiment.

\[
\begin{align*}
\frac{p_l + p_m + p_n}{p_l + 2p_m + 2p_n} & \quad \text{Any outcome where learner looks at red bottle} \\
\frac{p_m}{p_l + 2p_m + 2p_n} & \quad \text{Additional two outcomes where learner looks at other bottle}
\end{align*}
\]

\[
\begin{align*}
p_l &= p_m \frac{m}{m + n} + p_n \\
p_m &= p_m \frac{m}{m + n} \left(1 - p_l\right) \\
p_n &= (1 - p_m)(1 - p_l)
\end{align*}
\]

Category $N'$, antecedent = "red bottle"

Category $N$, antecedent = "bottle"

Category $N'$, antecedent = "bottle"

Learner Results

Averages over 1000 simulations, standard deviations in parentheses.

<table>
<thead>
<tr>
<th></th>
<th>Baker</th>
<th>R&amp;G</th>
<th>P&amp;L’s ED</th>
<th>P&amp;M</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_l$</td>
<td>0.58</td>
<td>0.95</td>
<td>0.02</td>
<td>0.09</td>
</tr>
<tr>
<td>$p_m$</td>
<td>0.50</td>
<td>0.97</td>
<td>0.17</td>
<td>0.37</td>
</tr>
<tr>
<td>$p_l$</td>
<td>0.50</td>
<td>0.97</td>
<td>0.17</td>
<td>0.37</td>
</tr>
</tbody>
</table>

As previous studies found:

Traditional unambiguous data alone fails (Baker).

Leveraging Sem Syn ambiguous data succeeds (R&G, P&L).

Leveraging Syn ambiguous data in addition fails (P&L’s ED).

New result: Leveraging Unamb NP data (P&M) does not yield the correct representation in general ($p_h$ is low), but...
Learner Results
Averages over 1000 simulations, standard deviations in parentheses.

<table>
<thead>
<tr>
<th></th>
<th>Baker</th>
<th>R&amp;G</th>
<th>P&amp;L’s EO</th>
<th>P&amp;M</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_l$</td>
<td>0.50 (+0.01)</td>
<td>0.95 (+0.01)</td>
<td>0.02 (+0.01)</td>
<td>&gt;0.99 (+0.01)</td>
</tr>
<tr>
<td>$p_{f_l}$</td>
<td>0.50 (+0.01)</td>
<td>0.97 (+0.01)</td>
<td>0.17 (+0.02)</td>
<td>0.37 (+0.04)</td>
</tr>
<tr>
<td>$p$(LWF behavior)</td>
<td>0.53 (+0.01)</td>
<td>0.93 (+0.01)</td>
<td>0.50 (+0.01)</td>
<td>&gt;0.99 (+0.01)</td>
</tr>
<tr>
<td>$p$(correct representation when producing LWF behavior)</td>
<td>0.22 (+0.01)</td>
<td>0.92 (+0.01)</td>
<td>&lt;0.01 (+0.01)</td>
<td>&gt;0.99 (+0.01)</td>
</tr>
</tbody>
</table>

New result:
The probability of producing the LWF behavior with this incorrect representation is high.

How does this work?
If $p_l$ is high, then when a property is mentioned (like "red"), the learner believes that property is relevant - which means the referent must include that property (RED BOTTLE).

Learner Results
Averages over 1000 simulations, standard deviations in parentheses.

<table>
<thead>
<tr>
<th></th>
<th>Baker</th>
<th>R&amp;G</th>
<th>P&amp;L’s EO</th>
<th>P&amp;M</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_l$</td>
<td>0.58 (+0.01)</td>
<td>0.95 (+0.01)</td>
<td>0.02 (+0.01)</td>
<td>&gt;0.99 (+0.01)</td>
</tr>
<tr>
<td>$p_{f_l}$</td>
<td>0.58 (+0.01)</td>
<td>0.97 (+0.01)</td>
<td>0.17 (+0.02)</td>
<td>0.37 (+0.04)</td>
</tr>
<tr>
<td>$p$(LWF behavior)</td>
<td>0.53 (+0.01)</td>
<td>0.93 (+0.01)</td>
<td>0.50 (+0.01)</td>
<td>&gt;0.99 (+0.01)</td>
</tr>
<tr>
<td>$p$(correct representation when producing LWF behavior)</td>
<td>0.22 (+0.01)</td>
<td>0.92 (+0.01)</td>
<td>&lt;0.01 (+0.01)</td>
<td>&gt;0.99 (+0.01)</td>
</tr>
</tbody>
</table>

What this means:
LWF’s assumption that correct behavior indicates the child has the correct representation does not seem to hold.

Or does it?
When the child produces the correct behavior in the LWF experiment, the probability that the child has the correct representation when making that interpretation is very high, even if the probability for the correct representation in general (e.g., when there is no modifier present) is very low.

Update: LWF were not wrong about children’s representation when interpreting utterances like those in their experiment.

Also, the other learners behave as LWF expect:
When they show the correct behavior, they have the correct representation. When they show incorrect behavior, they have the incorrect representation.
Recap & Implications

- Children may be able to learn the correct interpretation for one in certain situations (such as the LWF experiment) by broadening the set of data they consider relevant.

- Just because children demonstrate that they have the correct interpretation some of the time does not mean they have the correct representation all of the time.

Recap & Implications

- While children must eventually learn the correct representation of one, they do not necessarily need to do so by 18 months.

- Instead, they may realize that one's category is N (rather than N') at some later point.

One possibility:

[F&al] Leverage the syntactic distribution of one with innate domain-general statistical learning, by using subtle domain-specific semantic distinctions that indicate syntactic category.

"ball with stripes"  "side of the road"
"one with dots"     "one of the river"
[modifiers]        [complements = conceptually evoked by head noun]
[head noun = N']   [head noun = N']

The Acquisition Trajectory

I want it.
I want one.
Another one!

Do you see him?
Do you see one?

Before 18 months:
Need domain-specific knowledge
Recognize that one is similar to other anaphoric elements (it, him, etc.).

How to get it?
Derive it by using innate domain-general statistical learning abilities to observe the distribution of one compared to these other elements.

The Acquisition Trajectory

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

Before 18 months:
Track how often a mentioned property is important for a referent to have.

How to get it?
Use innate domain-general statistical learning abilities to track this.
The Acquisition Trajectory

18 months:
Be able to assign the correct interpretation to utterances like those in the LWF experiment. (Know that one is N’ in these cases.)

“Look - a red bottle! Do you see another one?”

The Acquisition Trajectory

“ball with stripes”
“side of the road”
“one with dots”
“one of the river”

After 18 months:
Need domain-specific knowledge about subtle semantic distinctions that indicate syntactic category in order to leverage the syntactic distribution of one with innate domain-general statistical learning.

How?
May come from innate domain-specific knowledge (UG) about language.

Back to the bigger questions

- When induction problems exist, what does it take to solve them?
  - What indirect evidence is available? How might a child leverage this evidence?
  - Broader data sets that are identifiable via innate domain-general learning abilities may be additional sources of useful information.

Back to the bigger questions

- When induction problems exist, what does it take to solve them?
  - What learning biases can get the job done, given the available data?
    - Are they necessarily innate and domain-specific (UG)?
      - In this case study, the first step may not involve this kind of knowledge, although achieving the final adult knowledge state may.

<table>
<thead>
<tr>
<th>Stage I</th>
<th>18-month-old behavior</th>
<th>Stage II</th>
</tr>
</thead>
<tbody>
<tr>
<td>derived domain-specific knowledge</td>
<td>innate domain-general statistical learning</td>
<td>innate domain-specific knowledge?</td>
</tr>
<tr>
<td>innate domain-general statistical learning</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Back to the bigger questions

- How can the necessary learning biases inform us about how the acquisition process works?
  - Identify learning biases needed to achieve 18-month-old behavior
  - Identify knowledge state those biases suggest
  - Suggest a two stage acquisition process for learning anaphoric one

The big picture

- Indirect evidence does not necessarily mean indirect negative evidence - it can come from considering a broader pool of informative data
- Indirect evidence does not necessarily negate the need for learning biases (of whatever kind)
- Considering indirect evidence and its impact on acquisition can help define concrete proposals about what is necessarily innate and domain-specific, and thus what is in Universal Grammar
- Knowing the impact of the necessary learning biases on acquisition may also inform us about the acquisition trajectory

Thank you

Ben Mis
The members of CoLa Lab at UCI
Erika Webb
Vance Chung
UCI Computational Models of Language Learning Seminar 2010
The National Science Foundation, grant BCS-0943896