UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

How Far Can Indirect Evidence Take Us? Anaphoric One Revisited

Permalink

Journal

ISSN

Authors

Pearl, Lisa
Mis, Benjamin

Publication Date

2011

Peer reviewed

syntactic antecedent of one = "red bottle"
semantic referent of one = RED BOTTLE

The original proposal for learning anaphoric one required children to have innate domain-specific knowledge about the structure of language, as part of the child's Universal Grammar (Baker, 1978). However, more recent studies have suggested alternative solutions involving innate domain-general learning abilities coupled with *input restrictions* that arise from domain-specific learning constraints (Foraker et al., 2009; Pearl & Lidz, 2009; Regier & Gahl, 2004).

Here, we consider whether indirect evidence leveraged from a broader input set could lead children to the correct knowledge for anaphoric one. If so, we can then refine the current views on what is required for successful acquisition - and specifically, whether it is (i) domain-specific or domain-general, and (ii) innate or derivable.

We first discuss adult and child knowledge of anaphoric one, and then review previous proposals for how to learn this from the available input. We then motivate why a child might view a broader input set as informative for anaphoric one. Following this, we present an online Bayesian learner that uses this broader data set, and find that our learner is indeed capable of reproducing the behavior associated with correct knowledge of anaphoric one without imposing any domain-specific input restrictions. We conclude with discussion of what is required for acquisition success, and how this impacts the larger debate about Universal Grammar.

English anaphoric one

Adult knowledge

The adult representation of English anaphoric one has both a syntactic and semantic component. In order to interpret an utterance like (1), the listener must first identify the syntactic antecedent of one, i.e., what string one is replacing. In (1), adults interpret one's antecedent as "red bottle", so the utterance is equivalent to "Look - a red bottle! Oh, look - another *red bottle*!"

Then, the listener uses this syntactic antecedent to identify the semantic referent of one, e.g., what object in the world one is referring to. Given the syntactic antecedent "red bottle", adults interpret the referent of one as a bottle that is red (RED BOTTLE), as opposed to just any bottle (BOTTLE). Ac-

Induction problems in language acquisition

One of the most controversial claims in developmental and theoretical linguistics is that children learning their native language face an induction problem, sometimes called "Poverty of the Stimulus" (Chomsky, 1980; Crain, 1991). Simply put, this is the claim that the data in children's input are insufficient to identify the correct language knowledge as quickly as children seem to.

If this is true, then children must bring something to the language acquisition problem - and the nature of this "something" is often debated. Is it domain-specific or domain-general? Is it something derivable from prior experience or something necessarily innate? These questions are important, as induction problems in language acquisition are often used to motivate innate, domain-specific knowledge about language (Universal Grammar (Chomsky, 1965)).

The potential induction problem presented by English anaphoric one (1) has received considerable recent attention (e.g., Foraker, Regier, Khetarpal, Perfors, and Tenenbaum (2009); Lidz, Waxman, and Freedman (2003); Pearl and Lidz (2009); Regier and Gahl (2004)).

(1) Anaphoric one
Situation: Two red bottles are present.
Utterance: "Look - a red bottle! Oh, look - another one!"
Interpretation of one:

How Far Can Indirect Evidence Take Us? Anaphoric One Revisited

Lisa S. Pearl (lpearl@uci.edu)
Department of Cognitive Sciences
3151 Social Science Plaza
Irvine, CA 92697

Benjamin Mis (bmis@uci.edu)
Department of Cognitive Sciences
3151 Social Science Plaza
Irvine, CA USA

Abstract

A controversial claim in linguistics is that children face an induction problem, which is often used to motivate the need for Universal Grammar. English anaphoric one has been argued to present this kind of induction problem. While the original solution was that children have innate domain-specific knowledge about the structure of language, more recent studies have suggested alternative solutions involving domain-specific input restrictions coupled with domain-general learning abilities. We consider whether indirect evidence coming from a broader input set could obviate the need for such input restrictions. We present an online Bayesian learner that uses this broader input set, and discover it can indeed reproduce the correct learning behavior for anaphoric one, given child-directed speech. We discuss what is required for acquisition success, and how this impacts the larger debate about Universal Grammar.

Keywords: anaphoric one; acquisition; Bayesian learning; domain-general; domain-specific; indirect evidence; input restrictions; language; online probabilistic learning; poverty of the stimulus; Universal Grammar

cording to standard linguistic practice, the string "red bottle" has the structure in (2), while "a red bottle" has the structure in (3):

(2) $[_{N'}$ red $[_{N'} [_{N^0}$ bottle]]]
(3) $[_{NP}$ a $[_{N'}$ red $[_{N'} [_{N^0}$ bottle]]]]

The syntactic category N^0 can only contain nouns (e.g., "bottle"), and the category NP contains any noun phrase (e.g., "a red bottle"). The syntactic category N' is larger than N^0 but smaller than NP, and can contain both nouns (e.g., "bottle") and noun+modifier strings (e.g. "red bottle"). Note that the string "bottle" can be labeled both as syntactic category N' (4a) and syntactic category N^0 (4b).[1]

(4a) $[_{N'} [_{N^0}$ bottle]]
(4b) $[_{N^0}$ bottle]

Linguistic theory posits that anaphoric elements only have antecedents of the same syntactic category. Since one's antecedent can be "red bottle", then one should be category N' in these cases. Notably, if the syntactic category of one were instead N^0, one could not have "red bottle" as its antecedent; instead, it could only have noun-only strings like "bottle", and we could not get the interpretation that we do for (1).

One way to represent adult knowledge is (5):

(5) Adult anaphoric one knowledge in utterances like "Look - a red bottle! Do you see another one?"
(a) Syntactic structure: category N'
(b) Semantic referent: The mentioned property ("red") is relevant for determining the referent of one.

Child knowledge

Behavioral evidence from Lidz et al. (2003) (henceforth **LWF**) suggests that young children also have this same interpretation for utterances like (1). LWF examined the looking behavior of 18-month-olds when hearing an utterance like "Look, a red bottle! Do you see another one?". The 18-month-olds demonstrated a significant preference for looking at the bottle that was red (as compared to a bottle that was some other color). LWF interpreted this to mean that by 18 months, children have acquired the same representation for anaphoric one that adults have.

Learning anaphoric one

The learning problem

Learning the correct representation for anaphoric one is difficult because many anaphoric one data are ambiguous with respect to what syntactic category one is, even if children know that the choice is between N' and N^0. Moreover, as we saw in (2), sometimes there is more than one N' antecedent to choose from (e.g., "red bottle" vs. "bottle"), which means

that there is also ambiguity with respect to the semantic referent (e.g., RED BOTTLE vs. any BOTTLE). Examples (6) and (7) demonstrate two kinds of ambiguous data.

(6) Syntactic (**Syn**) Ambiguity
Situation: There are two bottles present.
Utterance: "Look, a bottle! Oh look - another one!"

(7) Semantic and Syntactic (**Sem-Syn**) Ambiguity
Situation: There are two red bottles present.
Utterance: "Look, a red bottle! Oh look - another one!"

Syn ambiguous data do not clearly indicate the category of one, even though the semantic referent is clear. In (6), the semantic referent must be BOTTLE since the antecedent can only be "bottle". But, is the syntactic structure $[_{N'} [_{N^0}$ bottle]] or just $[_{N^0}$ bottle]? Notably, if the child held the mistaken hypothesis that one was category N^0, this data point would not conflict with that hypothesis since it is compatible with the structure being $[_{N^0}$ bottle].

Sem-Syn ambiguous data are unclear about both the referent and the category of one. In (7), if the child held the mistaken hypothesis that the referent is simply BOTTLE (unlike the adult interpretation of RED BOTTLE), this would not be disproven by this data point - there is in fact another bottle present. This data point is ambiguous syntactically for the same reason Syn data like (6) are: if the referent is BOTTLE, then the antecedent is "bottle", which is either N^0 or N'.

Fortunately, there are some unambiguous data available like (8), but these require a very specific conjunction of situation and utterance.

(8) Unambiguous (**Unamb**) data
Situation: Both a red bottle and a purple bottle are present.
Utterance: "Look - a red bottle! There doesn't seem to be another one here, though."

In (8), if the child mistakenly believes the referent is just BOTTLE, then the antecedent of one is "bottle" and it's surprising that the speaker would claim there's not "another bottle here", since another bottle is clearly present. Thus, this data point unambiguously indicates that the property "red" is important, so the semantic referent is RED BOTTLE. The corresponding syntactic antecedent is "red bottle", which has the syntactic structure $[_{N'}$ red $[_{N'} [_{N^0}$ bottle]]] and indicates one's category is N'.

Unfortunately, unambiguous data comprise only a small fraction of children's input - LWF discovered that a mere 0.25% of child-directed anaphoric one utterances were unambiguous data. For this reason, the debate has arisen about how children might solve this acquisition problem as rapidly as they do.

Innate, domain-specific knowledge

An early proposal (Baker, 1978) (henceforth **Baker**) assumed that only unambiguous data were informative. Given the sparsity of these data, it was proposed that children possess domain-specific knowledge about the structure of language -

---

[1]We note that the actual labels themselves are immaterial. It is only relevant that these structural levels are distinguished in this way, i.e., that "red bottle" and "bottle" are the same label (N' here), while "bottle" can also be labeled with a smaller category label (N^0 here).

in particular, children innately know that anaphoric elements (like one) cannot be syntactic category N^0. Instead, children automatically rule out that possibility from their hypothesis space, and simply know that one is category N'.

Domain-general learning abilities + domain-specific input restrictions

Regier and Gahl (2004) (henceforth **R&G**) noted that Sem-Syn data like (7) could be leveraged to learn the correct representation for anaphoric one. Specifically, a learner with domain-general statistical learning abilities could track how often a property that was mentioned was important for the referent to have (e.g., when "red" was mentioned, was the referent just a BOTTLE or specifically a RED BOTTLE?). If the referent had that property (e.g., was a RED BOTTLE), this meant that the syntactic antecedent of one would be an N' string (e.g., "red bottle") and implicated one's category as N'. The R&G data set consisted of both unambiguous data and Sem-Syn ambiguous data, and their online Bayesian learner was able to learn the correct interpretation for anaphoric one. No innate, domain-specific knowledge was required.

Pearl and Lidz (2009) (henceforth **P&L**) noted that if the child had to learn the syntactic category of one, an equal-opportunity (**EO**) learner would view Syn ambiguous data like (6) as informative. Unfortunately, Syn ambiguous data far outnumber the Sem-Syn ambiguous and unambiguous data combined (about 20 to 1 in their corpus analysis), and in fact lead a probabilistic learner of the kind R&G propose to the wrong syntactic category for one (i.e., one=N^0). Thus, R&G's Bayesian learner would have to explicitly filter out the Syn ambiguous data. P&L suggested that this kind of filter is domain-specific, since it involves ignoring a specific kind of linguistic data, though they speculate how this restriction could be derived from domain-general learning preferences.

Foraker et al. (2009) (henceforth **F&al**) focused on identifying the syntactic category of one, applying an ideal Bayesian learner to the syntactic input alone. Their learner employed subtle conceptual knowledge to identify the likely syntactic category for one (specifically, a syntactic complement is "conceptually evoked by its head noun" and indicates the noun string is N^0, while a modifier is not and indicates the noun string is N'). While there were not many informative one data points in their data, their ideal learner was able to learn that one was category N'. Their learner required a domain-specific input restriction to syntactic data as well as domain-specific knowledge about the subtle distinction between complements and modifiers, and their implications for syntactic categories.

A broader view of informative data

Instead of restricting the input set, we consider expanding it beyond unambiguous (8), Sem-Syn ambiguous (7), and Syn ambiguous (6) data. Consider that there are other anaphoric elements in the language besides one, such as pronouns like it, him, her, etc. These pronouns are category NP, since they replace an entire noun phrase (NP) when they are used (9):

(9) "Look at the cute penguin. I want to hug it/him/her."
≈ "Look at the cute penguin. I want to hug *the cute penguin*."

Here, the antecedent of the pronoun it/him/her is the NP "the cute penguin":

(10) [$_{NP}$ the [$_{N'}$ cute [$_{N'}$ [$_{N^0}$ penguin]]]]

In fact, it turns out that one can also have an NP antecedent:

(11) "Look! A red bottle. I want one."
≈ "Look! A red bottle. I want *a red bottle*."

So, the issue of one's syntactic category only occurs when one is being used in a syntactic environment that indicates it is smaller than NP (examples of this <**NP** environment are in (1), (6), (7), and (8)). Notably, one shares some semantic and syntactic distribution properties with other pronouns.

Following R&G's idea, a learner could track how often a property mentioned in the potential antecedent (e.g., "red" in "a red bottle" in (11)) is important for the referent to have. Crucially, we can apply this not only to data points where one is <NP ((6) and (8)), but also to data points where pronouns are used anaphorically and in an NP syntactic environment ((9) and (11)). When the potential antecedent mentions a property and the pronoun is used as an NP, the antecedent is necessarily also an NP, and necessarily includes the mentioned property (e.g., "a red bottle"). Data points like (9) and (11) are thus unambiguous both syntactically (category=NP) and semantically (the referent must have the mentioned property). We will refer to them as unambiguous NP (**Unamb NP**) data points, and these are the additional data points our learner (henceforth the **P&M** learner) will learn from.

Like the R&G and P&L learners, our learner differs from the Baker learner by learning from data besides the unambiguous <NP data. However, our learner differs from the learners in R&G and P&L by learning from data containing anaphoric elements besides one.[2] Table 1 shows which learners use which data.

Table 1: Data sets used by learners.

Data type	Example	Learners
Unamb <NP	(8)	Baker, R&G, P&L's EO, P&M
Sem-Syn Ambig	(7)	R&G, P&L's EO, P&M
Syn Ambig	(6)	P&L's EO, P&M
Unamb NP	(9), (11)	P&M

Information in the data

Figure 1 represents the information dependencies in any data point where a pronoun is used anaphorically and there is a potential antecedent that has been mentioned recently.

---

[2]Our learner also differs from the F&al learner by leveraging both syntactic and semantic information, instead of just syntactic information.
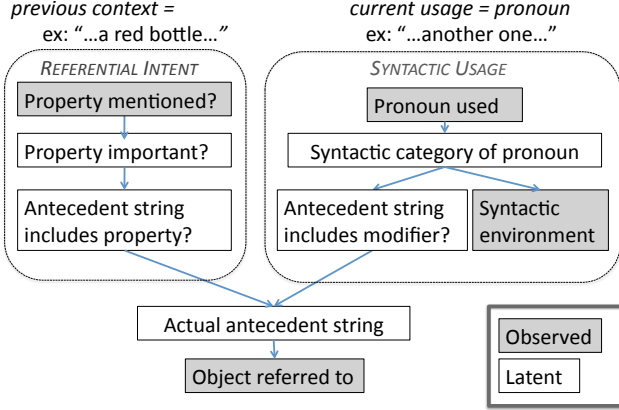
Figure 1: Information dependencies in data points.

Under SYNTACTIC USAGE, we can observe which pronoun is used (e.g., it, one, etc.). The syntactic category depends on which pronoun is used (e.g., NP, N', or N^0 for one). We can observe the syntactic environment in which the pronoun is used, which depends on the syntactic category (e.g., "another one" indicates a syntactic environment of $<$NP, which means the category is N' or N^0). The syntactic category also determines whether the antecedent string can contain a modifier (e.g., category N^0 cannot, since it only allows bare nouns like "bottle").

Under REFERENTIAL INTENT, we can observe whether the potential antecedent mentioned a property or not (e.g., "a red bottle" vs. "a bottle"). If a property was mentioned, it is a latent variable whether the mentioned property was important for the referent of the pronoun to have. This then determines whether the antecedent string must include that property (e.g., it must if the property is important).

Both the antecedent string variables determine the content of the actual antecedent string (e.g., if both a modifier and a property must be included, the antecedent would be "red bottle" rather than simply "bottle"). Finally, the antecedent string determines what object is being referred to, and whether that object has the mentioned property (e.g., whether it's a RED BOTTLE when the previous context was "a red bottle"). This is observable (e.g., we can see if the bottle that one refers to is in fact red).

These variables can take on the following values:

REFERENTIAL INTENT
property mentioned? $\in$ {Yes, No}
property important? $\in$ {Yes, No}
antecedent string includes property? $\in$ {Yes, No}

SYNTACTIC USAGE
pronoun used $\in$ {one, it, him, her, etc.}
syntactic category of pronoun $\in$ {NP, N', N^0}
syntactic environment $\in$ {NP, $<$NP}
antecedent string includes modifier? $\in$ {Yes, No}

COMBINED

actual antecedent string $\in$ {"red bottle", "bottle", etc.}
object referred to $\in$ {has property, does not have property}

The data types used by the different learners have the observable values in Table 2.

Table 2: Data types and observable values.

Data type	PropMent	Pronoun	SynEnv	Obj
Unamb $<$NP	Yes	one	$<$NP	has prop
Sem-Syn Ambig	Yes	one	$<$NP	has prop
Syn Ambig	No	one	$<$NP	N/A
Unamb NP	Yes	it, one, etc.	NP	has prop

The online probabilistic learning framework

Important quantities

The two components of the correct representation for anaphoric one are (a) that a property mentioned in the potential antecedent is important for the referent of one to have, and (b) that one is category N' when it is not an NP. These correspond to "property important?" and "syntactic category of pronoun" in Figure 1. We represent the probability of the former as p_I and the probability of the latter as $p_{N'}$.

We follow the update methods in P&L, and use equation (12) adapted from Chew (1971), which assumes p comes from a binomial distribution and the beta distribution is used to estimate the prior:

$$p_x = \frac{\alpha + data_x}{\alpha + \beta + totaldata_x}, \alpha = \beta = 1 \qquad (12)$$

α and β represent a very weak prior when set to 1. $data_x$ represents how many informative data points indicative of x have been observed, while $totaldata_x$ represents the total number of potential x data points observed. After every informative data point, $data_x$ and $totaldata_x$ are updated as in (13), and then p_x is updated using equation (12). The variable ϕ_x indicates the probability that the current data point is an example of an x data point. For unambiguous data, $\phi_x = 1$; for ambiguous data $\phi_X < 1$.

$$data_x = data_x + \phi_x \qquad (13a)$$
$$totaldata_x = totaldata_x + 1 \qquad (13b)$$

p_I is updated for Unambiguous $<$NP data, Sem-Syn Ambiguous data, and Unambiguous NP data. $p_{N'}$ is updated for Unambiguous $<$NP data, Sem-Syn Ambiguous data, and Syn Ambiguous data.

The value of ϕ_x depends on data type. We can derive the value of ϕ_I by using the information dependencies in Figure 1, and the basic Bayes equation. ϕ_I uses equation (14), which includes π (what pronoun was mentioned), σ (what the syntactic environment is), μ (whether the previous context mentioned a property), ω (whether the object has the mentioned property), and I (the property is important):

$$\phi_I = p(I|\pi,\sigma,\mu=yes,\omega) = \frac{p(\pi,\sigma,\omega|I,\mu=yes)*p_I}{p(\pi,\sigma,\omega|\mu=yes)} \quad (14)$$

Unambiguous <NP and Unambiguous NP data have $\phi_I=1$, which is intuitively satisfying since they unambiguously indicate that the property is important for the referent to have. Sem-Syn ambiguous data have ϕ_I calculated as in (15):

$$\phi_I = \frac{\rho_1}{\rho_1+\rho_2+\rho_3} \quad (15)$$

where

$$\rho_1 = \quad p_{N'} * \frac{m}{n+m} * p_I \quad (16a)$$
$$\rho_2 = \quad p_{N'} * \frac{n}{n+m} * (1-p_I) * \frac{1}{t} \quad (16b)$$
$$\rho_3 = \quad (1-p_{N'}) * (1-p_I) * \frac{1}{t} \quad (16c)$$

In (16), m and n refer to how often N' strings are observed to contain modifiers (m) (e.g., "red bottle), as opposed to containing only nouns (n) (e.g., "bottle"). These help determine the probability of observing an N' string with a modifier (16a), as compared to an N' string without one (16b). Parameter t indicates how many property types there are in the learner's hypothesis space, which determines how suspicious a coincidence it is that the object just happens to have the mentioned property.

The quantities in (16) correlate with anaphoric one representations. For ρ_1, the syntactic category is N' ($p_{N'}$), a modifier is used ($\frac{m}{n+m}$), and the property is important (p_I). For ρ_2, the syntactic category is N' ($p_{N'}$), a modifier is not used ($\frac{n}{n+m}$), the property is not important (1- p_I), and the object has the mentioned property by chance ($\frac{1}{t}$). For ρ_3, the syntactic category is N^0 (1-$p_{N'}$), the property is not important (1-p_I), and the object has the mentioned property by chance ($\frac{1}{t}$).

The value of $\phi_{N'}$ also depends on data type. We derive the value of $\phi_{N'}$ similarly (though not identically) to ϕ_I:

$$\phi_{N'} = p(N'|\pi,\sigma=<NP,\mu,\omega) = \frac{p(\pi,\mu,\omega|N',\sigma=<NP)*p_{N'}}{p(\pi,\mu,\omega|\sigma=<NP)} \quad (17)$$

Unambiguous <NP data have $\phi_I=1$, which is again intuitively satisfying since they unambiguously indicate that the category is N' when the syntactic environment is <NP. Sem-Syn ambiguous data have $\phi_{N'}$ as in (18):

$$\phi_{N'Sem-Syn} = \frac{\rho_1+\rho_2}{\rho_1+\rho_2+\rho_3} \quad (18)$$

where ρ_1, ρ_2, and ρ_3 are the same as in (16). Equation (18) is intuitively satisfying as only ρ_1 and ρ_2 are representations with syntactic category N'.

Syn Ambiguous data have $\phi_{N'}$ as the following:

$$\phi_{N'Syn} = \frac{\rho_4}{\rho_4+\rho_5} \quad (19)$$

where

$$\rho_4 = \quad p_{N'} * \frac{n}{n+m} \quad (20a)$$
$$\rho_5 = \quad 1-p_{N'} \quad (20b)$$

The quantities in (19) intuitively correspond to representations for anaphoric one when no property is mentioned in the previous context. For ρ_4, the syntactic category is N' ($p_{N'}$) and the N' string uses only a noun ($\frac{n}{n+m}$). For ρ_5, the syntactic category is N^0 (1-$p_{N'}$).

Learner input sets & parameter values

To gauge the frequency of the different data types in child-directed input, we conducted a corpus analysis of 17,521 child-directed utterances in the Brown-Eve corpus from CHILDES (MacWhinney, 2000). Following P&L, we posit that the anaphoric one learning period begins at 14 months and that children hear approximately 1,000,000 sentences from birth until 18 months. We can then use the data frequencies in the Brown-Eve corpus to estimate the expected distribution of pronoun data between 14 and 18 months. Table 3 shows the input sets used to test the different learning proposals for anaphoric one.

Table 3: Input sets for different anaphoric one proposals

Data type	Baker	R&G, P&L	P&L's EO	P&M
Unamb <NP	0	0	0	0
Syn-Sem Ambig	0	242	242	242
Syn Ambig	0	0	2743	2743
Unamb NP	0	0	0	3073
Uninformative	36500	36258	33515	30442

For the free parameters in the model, we will follow the corpus-based estimate P&L used for m and n: $m = 1$ and $n = 3$. We will also follow an estimate P&L used for t: $t = 5$.

Measures of success

In addition to measuring p_I and $p_{N'}$ at the end of the learning period, a good metric of acquisition success is how likely the learner is to produce the infant looking behavior in the LWF experiment (e.g., "Look - a red bottle! Do you see another one?"). Specifically, we can calculate the probability (p_{beh}) of the learner looking at the referent that has the mentioned property (e.g., the RED referent given that "red" was mentioned) when given a choice between two referents.

$$p_{beh} = p(\omega=hasproperty|\pi=one,\sigma=<NP,\mu=yes) \quad (21)$$

This works out to

$$p_{beh} = \frac{\rho_1+\rho_2+\rho_3}{\rho_1+2*\rho_2+2*\rho_3} \quad (22)$$

where ρ_1, ρ_2, and ρ_3 are defined as in (16), $m = 1$, $n = 3$, and $t = 2$. As before, these quantities intuitively correspond to the different outcomes. The numerator represents all the

outcomes where the learner looks to the correct object (ρ_1, ρ_2 and ρ_3 looking at the RED bottle), while the denominator includes the two additional outcomes where the learner looks to the incorrect object (ρ_2 and ρ_3 looking at the non-RED bottle).

Results & General discussion

Table 4 shows the results of the learning simulations over the different input sets, with averages over 1000 runs reported and standard deviations in parentheses.

Table 4: Probabilities after learning

Prob	Baker	R&G, P&L	P&L's EO	P&M
$p_{N'}$	.50 (<.01)	.97 (<.01)	.17 (.02)	.37 (.04)
p_I	.50 (<.01)	.95 (<.01)	.02 (.01)	>.99 (<.01)
p_{beh}	.53 (<.01)	.93 (<.01)	.50 (<.01)	>.99 (<.01)

Interestingly, while the P&M learner has a lower probability for one as N' in general ($p_{N'}$ = .37), it has an extremely high probability of reproducing infant behavior and interpreting one correctly in the LWF scenario (p_{beh} > .99). This is because the learner believes the mentioned property is important (p_I > .99). If the property is important, the antecedent must contain the modifier (e.g., be "red bottle" as opposed to "bottle") - which means the learner will choose the correct referent *even if the learner generally thinks one is N^0*. That is, infants could produce adult-like behavior in this context without having adult-like representations of anaphoric one.

We note that this result is due to the input set the P&M learner is using - the learners using restricted input sets behave exactly as previous studies found. Learning from unambiguous data alone does not work (Baker), though including Sem-Syn ambiguous data will lead to the correct representation and the correct behavior (R&G, P&L). Additionally including Syn ambiguous data (P&L's EO) leads to the incorrect representation and chance looking behavior. Expanding to unambiguous NP data (P&M) doesn't solve the incorrect category problem for one, but it turns out this isn't always necessary to interpret one correctly in context.

This suggests that while children must eventually learn that one is N', they do not need to do so by 18 months. This may allow them time to develop the ability to make the subtle conceptual distinctions F&al's learner uses to leverage the syntactic distribution of one and converge on one as N'. This leads to a more complex acquisition trajectory. Initially, children could use a broader input set (like the P&M learner) and learn the correct interpretation for one in most contexts, even if they believe one is usually N^0. Later, children could be sophisticated enough to leverage the information in the syntactic distribution and identify one as definitively N'.

The knowledge needed for acquisition success would then include both domain-specific and domain-general components. To identify the broader data set the P&M learner used, the child needs to recognize that one is similar to other pronouns (i.e., it is anaphoric and has syntactic antecedents).

This is domain-specific knowledge, though it could be derived through innate domain-general statistical learning abilities applied to the input. The child can then track how often a mentioned property is important for a referent to have by using these same domain-general abilities. In the second learning stage, the learner uses these domain-general abilities coupled with domain-specific knowledge that allows complements and modifiers to be distinguished in the syntactic input (F&al). This domain-specific knowledge could be innate (as part of Universal Grammar), or perhaps derived somehow.

To conclude, we find that indirect evidence can be leveraged effectively by an online probabilistic learner in order to produce behavior consistent with infant anaphoric one behavior, even if the learner does not achieve the adult representation. Though this learning step does not require innate domain-specific knowledge, a second step that allows the learner to achieve the adult representation might. We believe this general approach of looking at broader input sets for learning linguistic phenomena may be fruitful for identifying what is and is not necessarily part of Universal Grammar.

References

Baker, C. L. (1978). *Introduction to generative transformational syntax*. Englewood Cliffs, NJ: Prentice Hall.

Chew, V. (1971). Point estimation of the parameter of the binomial distribution. *American Statistician*, *25*(5), 47–50.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge: The MIT Press.

Chomsky, N. (1980). *Rules and representations*. Oxford: Basil Blackwell.

Crain, S. (1991). Language acquisition in the absence of experience. *Behavioral and Brain Sciences*, *14*, 597–612.

Foraker, S., Regier, T., Khetarpal, N., Perfors, A., & Tenenbaum, J. (2009). Indirect evidence and the poverty of the stimulus: The case of anaphoric one. *Cognitive Science*, *33*, 287–300.

Lidz, J., Waxman, S., & Freedman, J. (2003). What infants know about syntax but couldn't have learned: experimental evidence for syntactic structure at 18 months. *Cognition*, *89*, B65–B73.

MacWhinney, B. (2000). *The childes project: Tools for analyzing talk*. Mahwah, NJ: Lawrence Erlbaum Associates.

Pearl, L., & Lidz, J. (2009). When domain-general learning fails and when it succeeds: Identifying the contribution of domain-specificity. *Language Learning and Development*, *5*(4), 235–265.

Regier, T., & Gahl, S. (2004). Learning the unlearnable: The role of missing evidence. *Cognition*, *93*, 147–155.