Are infants sensitive to informant reliability in word learning?

Alayo Tripp, University of Minnesota
Naomi Feldman, University of Maryland
William Idsardi, University of Maryland

November 8, 2019
BUCLD 44

The question that i’m interested in is how social knowledge impacts linguistic representations, specifically in early word learning.
The Traditional Linguistic Model

- Social knowledge considered distinct from linguistic knowledge

- “Linguistic theory is concerned with an ideal speaker-listener in a completely homogeneous speech community, who knows its language perfectly.” (Chomsky, 1965)

- This abstraction is of limited utility - users of a given language are not homogenous!

- So, is infant word learning influenced by social beliefs about speech sources?

Traditionally, linguists have used models of language which abstract away from the complexity of social knowledge.

Chomsky famously defined linguistic theory as only concerned with speech of theoretical homogenous speech communities: in effect, the speech of people who are all identical. In reality, as language users, we only use language with people who aren’t identical to us.

So, is infant word learning influenced by social knowledge?
Of course it is! But...

- Infants know language users aren’t identical
- There’s a lot we don’t know about how they learn language variation
- Even though we know infants are engaged in social learning, **this kind of inference is often ignored in computational models of language learning**

**Of course it is.** This has long been known, and there is a growing body of work investigating how particular social processes affect early word learning (summarized thoroughly in the publication cited here). However, the problem of exactly how infant social beliefs impact acquisition hasn’t been addressed directly in models of infant language acquisition. I am going to show you an example today of how we can approach this problem.
Modeling source evaluation in word learning

- There’s an existing literature on this in older children
- We will apply an existing computational model to reinterpret the results of an infant word learning task (the Switch task)
- Developmental changes in source evaluation provide a simple explanation for differences in behavior between age groups

And that’s by drawing on the literature about source evaluation. There’s an existing literature on how older children selectively learn words. So I’m going to introduce to you a model of reasoning about source reliability introduced to capture the behavior of 4 and 5 year old children, and then show you that it can also explain the infant behavior on a test of word learning called the Switch task. In particular I’m going to show you simulations using this model of a variation on the Switch task published by Rost and McMurray in 2009. Ultimately I want to convince you that models of language must incorporate what we understand about how listeners engage in non-linguistic reasoning about informants, and that the literature on source evaluation, or epistemic trust is a really good place to start.

To begin I want to frame the problem of learning words in the context of variation.
A word learning problem

Suppose an infant wants to learn the name of this object.

Suppose we’re an infant and we want to learn the name of this object.
A word learning problem

Suppose we observe someone use this word form “daag” to label the object. To form an association between this word form and this meaning, and learn a label for this object, the infant will need to create a representation of the word form.
So for this simple example, where we have seen the object labeled once by one person, we could conclude that label daag that they gave is the right representation to have for this word. But what if we had more evidence?
So here we have four more people labeling the object, but only three of the five called it a daag. The speaker in blue called it a “dug” and the speaker in red called it a “dawg.” Supposing that all of these speakers are intentionally labeling this object, how can we decide what representation we should have for the label?

One way we might do it by filtering out who is a reliable informant for the specific speech variety they are acquiring. But, critically,
Standard models of word learning

infants expect all speakers to be equally useful linguistic informants.

(Pinker, 1979; Fried and Holyoak, 1984; De Boer and Kuhl, 2003; Chater and Manning, 2006; Xu and Tenenbaum, 2007; Norris and McQueen, 2008; Frank, Goodman, and Tenenbaum, 2009; McMurray, Aslin, and Toscano, 2009; Rasanen, 2012; Pajak, Bicknell, and Levy, 2013)

Traditional models of language acquisition operate under a simplifying assumption that the infant treats all sources of speech as equally reliable.
A word learning problem

Under a model like this, the infant will have to account for all this variation in labels using a single representation that effectively averages all the variation in the input. So our name for this object would be 60% daag, 20% dug and 20% dog.
I am going to try to convince you that in fact, infants do not learn an average of the input. Instead, they know that there are different language varieties, and only some speakers actually are speaking that same target speech variety that the infant is acquiring.
So this account predicts that infants will need to select one label that corresponds to their target speech variety, which is the correct pronunciation of the label, and account for other variation in how the object is labeled as coming from non-target speech varieties (here marked in orange).
So next I am going to review a bit of what we know about how older children evaluate informants for reliability on word learning tasks.
Informant reliability and word learning

4 and 5 year-old children

- Track the reliability of informants when learning labels (e.g. Corriveau and Harris 2009)
- Treat consensus as evidence of reliability (e.g. Corriveau, Fusaro and Harris 2009)
- Can be effectively modeled as using rational Bayesian inference to jointly determine both the label and the knowledgeability of the informant (Shafto, Eaves, Navarro and Perfors 2012)

Infants

- Track the reliability of informants in non-linguistic tasks (e.g. Chow et al 2008)

There’s a significant amount of literature showing that 4 and 5 yo children reason about the reliability of labelers when choosing who to learn words from.

In particular, when children have observed informants agreeing on labels with a group, they prefer to learn labels from THOSE informants, as opposed to from informants who they have seen dissenting from this consensus.

Pat Shafto, Baxter Eaves, Dani Navarro and Amy Perfors created a Bayesian model which modeled the children in these experiments as using selective trust to decide which informants to learn words from.

However, if we look to the infant literature most evidence for how infants evaluate informants comes from non-linguistic domains.
Do infants similarly infer the reliability of linguistic sources?

So what we’re going to do is evaluate whether infant behavior can be explained the same way that 4 and 5 year old’s behavior can be. I’m going to focus on 14-month olds in particular because word learning is well studied in this age group.
Now I'll describe this existing model of epistemic trust in word learning, which again, was created to describe the behavior of four and five year olds.
We begin with $C$, which represents the correct pronunciation of the label in the learner’s target speech variety. We suppose that there are three possible labels for the object, (dag, dug, dog) and each is equally likely to be the right one.
Modeling the word learning inference

K represents the speaker's knowledgeability - the speaker may either be a knowledgeable speaker of the target dialect (K=1) or not (K=0).

(Shafto, Eaves, Navarro and Perfors 2012)
Modeling the word learning inference

(Shafto, Eaves, Navarro and Perfors 2012)

Depending on what the correct pronunciation is in the learner’s dialect (C) and whether or not the speaker knows the correct pronunciation (K), the speaker then selects an intention (I) to articulate one of the possible labels.
Finally the speaker articulates the utterance, producing an acoustic token which we label D. In this case, an instance of “daag”
Modeling the word learning inference

We are making the simplifying assumption that knowledgeable speakers are guaranteed to produce correct intentions - so if the correct label is daag then a knowledgeable speaker will never intend to say otherwise.

(Shafto, Eaves, Navarro and Perfors 2012)
Modeling the word learning inference

We suppose that unknowledgeable speakers are also consistent, but more likely to use wrong labels so for our purposes, an unknowledgeable speaker selected at chance will have a one in three chance of being a good labeler despite their ignorance. So just one third of unknowledgeable speakers will correctly intend to call this a “daag” while another ⅓ of the unknowledgeable speakers will intend to label it as a dawg, and the remainder intend to label it a dug

(Shafto, Eaves, Navarro and Perfors 2012)
We can compare this model to the traditional story, where infants treat all speakers as equally useful, by simply assuming that all speakers are knowledgeable.
Inferring the category without source evaluation

The SINGLE speaker case

\[ P(C, D=\text{“daːg”}) \propto P(D=\text{“daːg”}|C) \cdot P(C) \]

| \( P(C=\text{daːg}) \) | 1 |
| \( P(C=\text{dɔːg}) \) | 0 |
| \( P(C=\text{dʌg}) \) | 0 |

And in this case, if we observe the label once, we should conclude that this is certainly the correct label.
Inferring the category without source evaluation

The MULTIPLE speaker case

And hearing the label a second time from the same source OR from a different source will not change our conclusion

\[
P(C,|D_1=\text{“da\:g”},D_2=\text{“da\:g”}) \propto P(D_1=\text{“da\:g”},D_2=\text{“da\:g”}|C) \ P(C)
\]

| P(C=da\:g)   | 1 |
| P(C=do\:g)   | 0 |
| P(C=d\:g)    | 0 |
Infants infer $P(K,C|D)$

$P(K=1)=0.5$
$P(K=0)=0.5$

$P(C=1)=1/3$  /daːɡ/  
$P(C=2)=1/3$  /dʰɑːɡ/  
$P(C=3)=1/3$  /dəːɡ/

(Shafto, Eaves, Navarro and Perfors 2012)

However, in the model created by shafto et al the infant needs not only to infer what the correct pronunciation of the label is in their target dialect, but also whether the speaker providing that label is in fact a speaker of that dialect or not. And we can walk through the predictions of this model to see how they are different.
Jointly inferring category and informant quality

The SINGLE speaker case

Supposing we hear one label from one source, now instead of three hypotheses, we now have six, because in addition three options for the label C, we also have two options for K.
Jointly inferring category and informant quality

The SINGLE speaker case

So following our assumptions, we would estimate that after observing this label once, we should believe that the label probably is dag, (as indicated by the sum of the green cells here, 2/3) but because we had a prior belief that the speaker was equally likely to be knowledgeable as not, we end up with a belief that with $\frac{1}{3}$ probability the label is actually NOT dag at all (this probability that the given label is incorrect is given by the sum of the orange cells)
Jointly inferring category and informant quality

The MULTIPLE speaker case

$$P(C, K_1, K_2 | D_1 = \text{da:g}, D_2 = \text{da:g}) \propto P(D_1 = \text{da:g}, D_2 = \text{da:g} | C, K_1, K_2) P(C, K_1, K_2)$$

<table>
<thead>
<tr>
<th>C=da:ɡ</th>
<th>K_1=1, K_2=1</th>
<th>K_1=1, K_2=0</th>
<th>K_1=0, K_2=1</th>
<th>K_1=0, K_2=0</th>
</tr>
</thead>
<tbody>
<tr>
<td>C=da:ɡ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C=da:ɡ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C=da:ɡ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

But if we hear a second label from a second source, in this case, instead of 6 possibilities, our hypothesis space now contains 12 possibilities, accounting for the fact that either, the first, the second, neither or both speakers might be knowledgeable.
Jointly inferring category and informant quality

The MULTIPLE speaker case

In this case, we should be more confident about the label - the probability that the label is correct (that’s the green cells) has now grown to 8/9, and the probability that the label is wrong (that’s the orange cells) has shrunk to 1/9

In fact, as the number of speakers grows, we can intuit that
First: the size of the hypothesis space will grow exponentially
Second: The probability assigned to the case where ALL of the observed speakers are unknowledgeable will continue to shrink

So basically, the more people we see testify to a label, the more certain we can be that at least one of them knows what they are talking about
So we’ve seen that these two different models make very different predictions about how listeners will interpret increasing amounts of testimony. Now I am going to describe the Switch task to you, and then demonstrate that the model I have introduced better simulates 14-month old behavior on this test of word learning than a model which assumes all speakers are equally useful.
The Switch Task

Violation of expectation paradigm (Stager and Werker 1997)

In this experiment, the infant is brought into the lab and situated in front of a screen which displays a novel object. They then hear a voice repeat a label for this object. Once the infant has habituated, they receive one of two types of test trials - same trials where they hear the same label again, and switch trials where they hear a different label. (the labels i’m using here as examples were not the labels originally used in that task but they were used in the reproduction of that task I am going to simulate for you)

So if the infant has learned to associate the label they are hearing with this object during the exposure phase, then the switch trials where they label changes should be surprising to them, and they should look longer for switch trials, than they look when the label stays the same. They predicted that the infants would ALWAYS LOOK LONGER TO SWITCH TRIALS
On the y-axis here we have the infants looking time, or how much time they spent looking at the screen. The significantly longer gray bar on the left shows that 8 mo old infants did spend more time looking at the object when the label was switched. But the difference between the bars on the right is not significant, showing that the 14-month old infants did not look longer during switch trials, when the label changed. So the prediction that infants will look longer to the switch trials does describe the behavior of 8 month olds, but it does not describe the behavior of 14 month olds.
The Switch Task

Violation of expectation paradigm (Stager and Werker 1997)

Habituation

Same Trial

Switch Trial

To ensure that this result was an effect of the infants learning to associate the object and the label, and not a purely an effect of their auditory abilities Stager and Werker ran another version of the experiment, and this time instead of seeing a novel object on the screen, the infants saw a checkerboard pattern. This pattern was chosen because it’s visually interesting, but unlikely to be interpreted as a labelable object.

Again, they heard the same label repeatedly, and then in test either heard the same label or they heard a switch trial, where the label was changed.
Now in the context of the checkerboard, the 14-month old behavior now looked like the 8 month olds, who attended to the switch trials, looking longer when the label changed than when it stayed the same. They concluded that the inability of the 14 month olds to detect the switch specifically in the presence of a labeled object (as opposed to in the checkerboard condition) demonstrates that the Switch task does test their word learning behavior.

I am going to show you a follow up study that exposed the infants to multiple speakers in the switch task, and then I’m going to use the two models we compared earlier to simulate those experiments.
Rost and McMurray hypothesized that the reason 14-month olds fail at this word learning task is because they have immature phonetic representations. They reasoned that if the infants were exposed to more variation in the input, they would better be able to learn the label. So in the first condition, they replicated the switch task, showing the infants the same label repeatedly from a single informant
And they successfully replicated Stager and Werker’s results, showing that in the context of learning a label for a novel object, the 14-month olds did not look longer when the label changed than when it stayed the same.
However, in a second condition, instead of providing the infants with a label being repeated by a single speaker, the infants were exposed to tokens of the label from different speakers. Again, they were tested in two conditions: the same trial, where the label was identical to the exposure phase, and the switch trial, where the label changed
When the infants were exposed to the label from multiple speakers, then they did look longer to the switch trial, suggesting that they had learned the label in more phonetic detail. Rost and McMurray attributed this to the increased variability in the acoustic information the multiple speakers provided, which could compensate for the infants’ immature phonetic representations, and help them learn the label.
Now that I have shown you this evidence of how 14-month olds use of phonetic representations when they are or aren’t doing word learning and how this can be explained by assuming that their phonetic representations are immature, I want to demonstrate that we can still predict this pattern of results while assuming that that infants’ phonetic perception. I am going to argue that the presentation of labels from multiple speakers may support a social inference about the quality of the informant.
Standard models of word learning

(Pinker, 1979; Fried and Holyoak, 1984; De Boer and Kuhl, 2003; Chater and Manning, 2006; Xu and Tenenbaum, 2007; Norris and McQueen, 2008; Frank, Goodman, and Tenenbaum, 2009; McMurray, Aslin, and Toscano, 2009; Rasanen, 2012; Pajak, Bicknell, and Levy, 2013)

infants expect all speakers of their language to be equally useful linguistic informants.

So the traditional model predicts that infants will treat all language sources as equally useful, and we can walk through its predictions about the pattern of looking in the switch task.
Assuming that all speakers are equally useful, we can just set $K=1$ for all informants.
Inferring the category without source evaluation
The SINGLE speaker case

\[ P(C|D=\text{“da:g”}) \propto P(D=\text{“da:g”}|C) P(C) \]

<table>
<thead>
<tr>
<th>C</th>
<th>1</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>C=buk</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>C=puk</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>C=duk</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

In this case, if we see the label buk once from a single informant, we should conclude that the label is buk
Inferring the category without source evaluation

The MULTIPLE speaker case

\[ P(C|D_1=\text{d\alpha:g}, D_2=\text{d\alpha:g}) \propto P(D_1=\text{d\alpha:g}, D_2=\text{d\alpha:g}|C) P(C) \]

<table>
<thead>
<tr>
<th></th>
<th>C=buk</th>
<th>C=puk</th>
<th>C=duk</th>
</tr>
</thead>
<tbody>
<tr>
<td>C=buk</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C=puk</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C=duk</td>
<td>0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

If we see the label a second time from a second informant, this extra evidence isn’t expected to impact our conclusion.
Inferring the category without source evaluation

**The SINGLE speaker case**

\[
P(C|D_1=D_2=D_3=D_4=D_5=D_6=D_7=\text{“buk”}) \propto P(D_1=D_2=D_3=D_4=D_5=D_6=D_7=\text{“buk”}|C) P(C)
\]

<table>
<thead>
<tr>
<th>C</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>C=buk</td>
<td>1</td>
</tr>
<tr>
<td>C=puk</td>
<td>0</td>
</tr>
<tr>
<td>C=duk</td>
<td>0</td>
</tr>
</tbody>
</table>

Similarly, if we saw the label seven times from one informant or
Inferring the category without source evaluation

The MULTIPLE speaker case

\[ P(C|D_1=D_2=D_3=D_4=D_5=D_6=D_7=\text{“buk”}) \propto P(D_1=D_2=D_3=D_4=D_5=D_6=D_7=\text{“buk”}|C) \ P(C) \]

<table>
<thead>
<tr>
<th>C</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>buk</td>
<td>1</td>
</tr>
<tr>
<td>puk</td>
<td>0</td>
</tr>
<tr>
<td>duk</td>
<td>0</td>
</tr>
</tbody>
</table>

Seven different times from seven different informants, we would still expect the same result
The traditional model predicts that the switch trials will always be more surprising than the same trials, and so looking to the switch trial should always be longer than looking to the same trial. This does predict how the 8-month olds behave, and how the 14-month olds behave when they are not learning the name of an object. But for the 14-month olds who are doing word learning, this model makes the wrong prediction.

However, i’m going to show you that the social model, which assumes that infants are uncertain about whether speakers are knowledgeable or not, can make the right predictions.
Applying the Shafto et al 2012 model

So in the source selection model, rather than just inferring the correct pronunciation, in this model the learner also needs to determine whether the speaker who provided the label is knowledgeable about the target dialect or not.
Inferring the category using source evaluation

The SINGLE speaker case

\[
P(C,K|D_1=D_2=D_3=D_4=D_5=D_6=D_7=\text{“buk”}) \propto P(D_1=D_2=D_3=D_4=D_5=D_6=D_7=\text{“buk”}|C,K) P(C,K)
\]

<table>
<thead>
<tr>
<th></th>
<th>K=1</th>
<th>K=0</th>
</tr>
</thead>
<tbody>
<tr>
<td>C=buk</td>
<td>(\frac{1}{2})</td>
<td>(\frac{1}{6})</td>
</tr>
<tr>
<td>C=puk</td>
<td>0</td>
<td>(\frac{1}{6})</td>
</tr>
<tr>
<td>C=duk</td>
<td>0</td>
<td>(\frac{1}{6})</td>
</tr>
</tbody>
</table>

For the single speaker case, we estimate that the probability that the infant believes the speaker is unknowledgable after hearing them repeat the label seven times is estimated to be \(\frac{1}{3}\)

(We only need to infer one K here, for the one speaker, so we can just compute this analytically)
Inferring the category using source evaluation

The MULTIPLE speaker case

\[ P(C,K_1,K_2,K_3,K_4,K_5,K_6,K_7,D_1=D_2=D_3=D_4=D_5=D_6=D_7=\text{"buk"}) \propto \prod_i P(D_i|C,K_i) P(C,K_i) \]

| | $|K=1|>0$ | $|K=1|=0$ |
|---|---|---|
| C=buk | 0.9988 | .0004 |
| C=puk | 0 | .0004 |
| C=duk | 0 | .0004 |

However, we predict that if the infant hears the label from seven different people, the probability that they will believe that all the informants they have observed so far are unknowledgeable is much smaller

(In this case instead of inferring one K, we needed to infer seven different Ks. Because of the complexity, we estimated this using Gibb’s sampling.)

Again, intuitively, the more people testify to a particular label, the more likely it is that at least one of them knows what they are talking about
So while the traditional model predicts that the infants will behave the same in these two conditions, the model I’ve shown you accurately predicts that infants will look longer to the switch trial in the multiple speaker condition.
Standard models of word learning

infants expect all speakers of their language to be equally useful linguistic informants.

Social model of word learning

infants know that some speakers of their language are better models than others.

This evidence is consistent with the hypothesis that infants rely on inferences about the knowledgeability of individual speakers when deciding who to learn words from.
Discussion

• The Shafto et. al (2012) model reproduces the pattern of results from Rost and McMurray (2009).

• This model was previously applied to explain the labeling behavior of 4-5 year old children.

• It also accurately models the behavior of 14 month old infants

• Previous explanations characterized 14-month old infants as having immature phonetic representations, but this model treats their phonetic knowledge as adult-like, and explains differences between younger and older infants as the result of social learning.

This model was originally created to describe how 4 and 5 year olds selectively trust informants to learn words from, but i’ve shown that it also captures the behavior of 14 month olds in a word learning task

The Rost and McMurray explanation for the 14-month old behavior on the Switch task posited that infants have immature phonetic representations, but i’ve shown that we can predict infant looking behavior while assuming that infant phonetic knowledge is adult like

so next I want to talk about some other manipulations to the Switch task that also been shown to improve 14-month old performance, and how this model can also accommodate those explanations
Explanations for the Switch Task

- Infants fail at the task because it is too cognitively demanding

What makes this task “cognitively challenging” is inferring informant quality

(Yoshida, Fennel, Swingley and Werker 2009; Fennell and Werker 2003; Fennell et al. 2007)

One explanation that’s been suggested for why 14-month olds fail on the Switch task is that the task is too difficult. What i’ve shown here is a specific problem which makes the task difficult - deciding which informants to trust
Explanations for the Switch Task

- Infants fail at the task because they aren’t sure that the speech act is intended to label the object.
- Improved performance with:
  - Familiar objects and labels
  - Use of sentence frames
  - Pretest training demonstrating labeling behavior

Measures that reduce referential ambiguity also suggest the informant will be of high quality

(Fennell and Werker 2003; Fennell and Waxman 2006; Fennell and Waxman 2010)

Another explanation that has a lot of support is that infants don’t attend to the switched label because it’s not clear that the label was actually intended to refer to the object. Modifications to the task which emphasize the relationship between the utterance and the object have been shown to improve performance, but I argue that these changes to the task also license the inference that the speaker is knowledgeable.
Conclusions & Implications

- **Non-linguistic source evaluation** provides a simple and useful way to investigate the effect of social knowledge on infant word learning.

- This model can potentially be used to (re)interpret other results from infant speech perception.

- Developmental differences in social knowledge could matter a lot for understanding how different groups of children learn language.

So to sum up, non-linguistic source evaluation gives us a foundational knowledge base that we can use to investigate the acquisition of fine phonetic detail. We know social processes impact the perception of phonetic detail, so it’s critical that we investigate what typical development of sociolinguistic perception looks like. Understanding what constitutes normal variation has the potential to improve both how we design and interpret clinical and educational assessments of language ability.
Thank you!

Jeff Lidz
Jan Edwards
Yi Ting Huang
Alexander Shushunov
Ben Munson
The Probabilistic Modeling Lab at the University of Maryland

This work funded by
NSF grants BCS-1320410 and IIS-1421695