Papafragou & Musolino (2003)
Huang & Snedeker (2009b)
Barner, Brooks, & Bale (2011)

Mike Frank
Psych/Ling 236
4/30/12
General motivations for devo work

• Understanding pragmatic development
  – Children are great at learning words by thinking about other people
  – But terrible (?) at making inferences about what people say

• Understanding the nature of scalar inferences
  – Perhaps differences in development will reveal important theoretical differences in mechanism
Pragmatics: A developmental puzzle

- Children are experts at pragmatic word learning, but terrible at scalar implicature
- Pragmatic word learning requires intention reading (Bloom, 2002), knowledge of discourse status (Akhtar et al., 1996), and some type of perspective taking (Baron-Cohen, Baldwin, & Crowson, 1997)
- All of these are Grice-relevant abilities; what’s missing?
Social learning

• **Coincide**: child looks at A, exp. names A
• **Conflict**: child looks at B, exp. names A

Children represent speakers’ intentions!

Baldwin (1993)
Disambiguation/mutual exclusivity

- What causes this kind of inference?
- Lexical principle of mutual exclusivity (Markman)
- Pragmatic inference (Clark)
  - Conventional form: “ball”
  - Contrast in meaning
- Note inferential similarity to scalar implicature

Markman & Wachtel (1988)
Searching for dissociations

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Underspecification</td>
<td>Said</td>
<td>Said</td>
<td>Explicature</td>
<td>Impliciture</td>
<td>Presumptive</td>
</tr>
<tr>
<td>Numeral upper-bounds</td>
<td>GCI</td>
<td>Said</td>
<td>Explicature</td>
<td>Impliciture</td>
<td>Presumptive</td>
</tr>
<tr>
<td>Ordering with <em>and</em></td>
<td>GCI</td>
<td>CI</td>
<td>Explicature</td>
<td>Impliciture</td>
<td>Presumptive</td>
</tr>
<tr>
<td>Lexical scalar inference</td>
<td>GCI</td>
<td>CI</td>
<td>Explicature</td>
<td>Impliciture</td>
<td>Presumptive</td>
</tr>
<tr>
<td>Particularized inferences</td>
<td>CI</td>
<td>CI</td>
<td>CI</td>
<td>CI</td>
<td>CI</td>
</tr>
</tbody>
</table>

- The project of distinguishing these theoretical classes is finding phenomena that dissociate them
  - E.g. numerals are fast, lexical SIs are slow (Huang & Snedeker, 2009a)

- Differences in processing and development, as well as linguistic properties, can provide dissociations
Some of the horses jumped over the log

some / two / started
In the analysis below, our dependent measure is the proportion of ‘No’ responses to the puppet’s statements, i.e. the subjects’ tendency to judge these statements as ‘bad’ descriptions of the stories they witnessed. The proportions of ‘No’ responses were entered into an analysis of variance (ANOVA) with two factors: age (5-year-olds vs. adults) and scale type (all, some, three, two). The analysis revealed a significant main effect of age ($F(1; 54) = 135.34, P < 0.0001$), a significant main effect of scale type ($F(2; 54) = 13.03, P < 0.0001$) and a reliable interaction between age and scale type ($F(2; 54) = 7.43, P = 0.001$) (see Fig. 3).

In test trials, we found that adult subjects overwhelmingly rejected the puppet’s statements in each of the three conditions, i.e. 92.5% of the time in the all, some condition, 100% of the time in the three, two condition and 92.5% of the time in the finish, start condition. Statistical analysis revealed no reliable difference between these rejection rates ($F(2; 27) = 1.92, P = 0.16$). By contrast, we found that while 5-year-olds rejected the puppet’s statements in the case of three, two 65% of the time, they almost never did so in the case of all, some and finish, start (12.5% and 10% of the time, respectively). This difference was confirmed statistically ($F(2; 27) = 11.17, P < 0.001$). Pairwise comparisons (Tukey–Kramer) further revealed a reliable difference between three, two – all, some and three, two – finish, start ($P = 0.002$ and $P = 0.001$, respectively) but no reliable difference between all, some and finish, start ($P = 0.77$).

On the control items, adults gave correct responses 100% of the time in the all, some condition, 80% of the time in the three, two condition and 95% of the time in the finish, start condition. No reliable difference was found between these means ($F(2; 27) = 2.43, P = 0.1$). On the same items, children gave correct responses 90% of the time in the all, some condition, 95% of the time in the three, two condition and 85% of the time in the finish, start condition. Specifically, five children rejected the puppet’s statements on all four of the test trials, one child on three of the test trials, one child on two of the test trials, one child on one of the test trials and two children rejected none of the test trials. In sum, six of the ten children almost always rejected the puppet’s statements (i.e. on three or four of the test trials), three children almost never rejected the puppet’s statements (i.e. on either none or one of the test trials) and one child rejected half of the test trials and accepted the other half.
Papafragou & Musolino (2003) E2

• TVJT focused on pragmatic felicity rather than strict truth
  – May not have been obvious to children
• Modify paradigm to make it more obvious
  – Training on puppet who says infelicitous things
  – All situations competitions, puppet comments on the result of the competition
some in Experiment 2 is that such statements give rise to the inference not all (which the children correctly reject). It is hard to see what other reason might be driving children’s rejections. It is highly unlikely that children could detect infelicity in such specific ways without having derived the scalar implicature.

One further fact which makes it clear that children are computing scalar inferences is that their performance is subject to contextual factors which are known to affect scalar implicatures (such as the level of cognitive expectations built into the conversational exchange). Of our two experiments, the second one invites scalar inferences more readily because of the presence of salient contextual cues which raise the expected level of cognitive effects. This is consistent with the adult intuition that the implicature is ‘clearer’ in Experiment 2. We find that in this experiment, children’s performance is significantly better than in Experiment 1 (even though the overall shape of results remains the same). In sum, we conclude that detecting the weakness of a statement (e.g. some vs. not all) and correcting it in accordance with cognitive expectations of informativeness is a good indication of the ability to spontaneously derive a scalar implicature.

6. General discussion

The experiments presented here explore young children’s ability to derive pragmatic inferences during utterance comprehension. Our specific target was a well known group of pragmatic inferences, scalar implicatures, which arise whenever a ‘weaker’ proposition (e.g. Some men have beards) is used to communicate that a stronger proposition does not hold (Not all men have beards). We were mainly interested in whether young (but otherwise linguistically sophisticated) children compute such implicatures, as adult communicators routinely do. We also explored the role of (a) the semantics of specific scalar expressions, and (b) the context of the experimental scenarios, as well as the nature of the task, in the derivation of these conversational inferences.
Katsos & Bishop (2011) – brief interlude

Table 1
Proportion of type of response in experiment 2.

<table>
<thead>
<tr>
<th>Type of utterance</th>
<th>Type of response</th>
<th>Scalar</th>
<th>Non-scalar</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimal</td>
<td>3 – ‘huge’</td>
<td>85</td>
<td>100</td>
<td>92.5</td>
</tr>
<tr>
<td></td>
<td>2 – ‘big’</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1 – ‘small’</td>
<td>15</td>
<td>0</td>
<td>7.5</td>
</tr>
<tr>
<td>Underinformative</td>
<td>3 – ‘huge’</td>
<td>0</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>2 – ‘big’</td>
<td>89</td>
<td>85</td>
<td>87</td>
</tr>
<tr>
<td></td>
<td>1 – ‘small’</td>
<td>11</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>False</td>
<td>3 – ‘huge’</td>
<td>5</td>
<td>0</td>
<td>2.5</td>
</tr>
<tr>
<td></td>
<td>2 – ‘big’</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1 – ‘small’</td>
<td>95</td>
<td>100</td>
<td>97.5</td>
</tr>
</tbody>
</table>

Children detect underinformativeness when given the option (though this doesn’t mean they compute SIs)
Huang & Snedeker (2009b)

- Comparison of children with adults
- E1: some / all / two / three
- E2: will “some” always be slow? Or only in cases where implicature is necessary?
- E3: will children ever notice implicature inconsistent material?
In Experiment 1, examples of visual-world displays for (A) some, (B) two, (C) all, and (D) three are shown. These displays illustrate the concept of quantifiers and their impact on the participant's actions. The figure shows the proportion of switches to and off the target over time, indicating the disambiguation phase and end phase of the experiment. The graphs compare different quantifiers: 'some', 'two', 'three', and 'all', each with distinct patterns of switching behavior. The data suggests that children are able to process quantifiers and make appropriate selections based on the context provided. The figure also indicates that some quantifiers may lead to faster disambiguation than others, reflecting differences in how children process and interpret these terms.
One-referent trials and (B) two-referent trials. Participants were instructed (girl with no socks). In the two-referent trials, we again introduced two boys with three socks) and a fourth character received nothing on these trials, three characters evenly shared nine items (girl and pairs. contrasted a subset quantity of one item with its empty set. Partation of

The materials compared the interpre-

switches off the Target. Thus, we used a different anal-

fied restrict the referent of

2

15.83,

F

H9257

but no interaction between age and condition,

SEMANTIC MEANING AND PRAGMATIC INTERPRETATION

and occupation was never recorded, but information from the 2000 years 0 months) participated in this study. The children were

right]. (8) feet [experimenter places a blank card next to the girl on the lower

three socks next to the boy on the upper left, and three socks next to

couch. The coach gave socks to Judy and socks to Craig and socks to

in two different referential contexts (see Figure 5). In

Two-referent trials should be no different than in trials where

contrast, if they never calculated the inference, processing in these

should be delays when

adults in using the control quantifiers to restrict reference and

altogether fail to generate scalar implicatures during comprehen-

suggest that resolution of the Target is quicker via semantic

of switches to target

Quantifier Phase

Disambiguation Phase

End Phase

0.8

0.6

0.4

0.2

0

0 ms 200 ms 400 ms 600 ms 800 ms 1000 ms

Proportion of switches to target

Proportion of switches off target

Quantifier Phase

Disambiguation Phase

End Phase

0 ms 200 ms 400 ms 600 ms 800 ms 1000 ms

Proportion of switches off target

Proportion of switches to target
E3

**Method**

Target fixations in the SI-violating trials were compared to those in the SI-consistent trials. The proportion of Target looks did not differ significantly between conditions, suggesting that the failure in scalar implicature is not due to a lack of interest in the Target. However, during the disambiguation phase, we found that the proportion of Switches to the Target was higher in the SI-violating trials, indicating that children were more likely to look away from the Target when it was ambiguous.

**Results**

In adults, the presentation of the Target prior to the onset of the quantifier did not affect the proportion of Switches to the Target. This suggests that adults were able to calculate scalar implicatures. In contrast, children showed significant delays in looking at the Target when it was ambiguous, indicating that they were unable to calculate scalar implicatures.

**Discussion**

These results suggest that children are not capable of calculating scalar implicatures during online language processing. This is in contrast to adults, who are able to calculate scalar implicatures quickly. The failure to calculate scalar implicatures could be due to the pragmatic infelicity of the utterance. Children may not be able to calculate scalar implicatures because they are not as aware of the pragmatic implications of their utterances.

**Conclusion**

Our findings demonstrate that children as well as adults initially prefer the alternative interpretation of the Target when it is ambiguous. However, adults are able to resolve the ambiguity quickly, while children continue to look away from the Target, indicating that they are unable to calculate scalar implicatures.
Papafragou & Tantalou (2004)

- **Quantifier**
  - Q: “Did you eat the oranges?” Tiger: “I ate some.”
  - Does Tiger get the prize for eating all?
- **Encyclopedic (based on world knowledge)**
  - Q: “Did you eat the sandwich?” A: “I ate the cheese.”
- **Ad-hoc (based on specifics of situation)**
  - Q: “Did you wrap the gifts?” A: “I wrapped the parrot.”

<table>
<thead>
<tr>
<th>Condition</th>
<th>Test trials</th>
<th>Control trials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantifier</td>
<td>77.5%</td>
<td>97.5%</td>
</tr>
<tr>
<td>Encyclopedic</td>
<td>70%</td>
<td>100%</td>
</tr>
<tr>
<td>Ad hoc</td>
<td>90%</td>
<td>92.5%</td>
</tr>
</tbody>
</table>
Barner, Brooks, & Bale (2011)

I. Compute basic meaning of a sentence $S$ containing $L$, a scalar item.
II. Generate a set of alternatives $(a_1, a_2, \ldots, a_n)$ to $S$, called $S_{alt}$. These are all the sentences that can be generated by replacing $L$ with its scalar alternatives.
III. Restrict the alternatives in $S_{alt}$ by removing any alternative that is entailed by the original utterance $S$. Call this restricted set $S^*$. 
IV. Strengthen the basic meaning of $S$ (containing $L$) with the negation of all of the members of $S^*$.

• Step 2 seems to be the problematic one, test this by A) providing alternatives and B) restricting them using only
• Failure modes:
  – Not knowing the alternatives
  – Not being able to hold them in mind (working memory)
Neither the word only nor the quantifier was emphasized by the experimenter's prosody.

3. Results

The use of the word only had a significant effect on how children interpreted sentences involving contextual alternatives, but had no effect on their interpretation of sentences involving context-independent alternatives (some and all).

First, consider the data for context-independent alternatives (see Fig. 2). In contexts where two of three items fit a description (e.g., two out of three animals are sleeping), children correctly agreed to sentences like, “Are some of the animals sleeping?” (2-Item True trials) on 80.0% of trials, and correctly denied that all of the animals were sleeping (2-Item False trials) on 87.2% of trials. On 2-Item True trials, use of the word only had no significant effect on children’s judgments (t(28) = 0.00, p > 0.05). Only was never used in 2-Item False trials, so it is not surprising that there was no difference between children’s responses in the only and no-only conditions (t(28) = 0.96, p > 0.05). Critically, on 3-Item Test trials, children in the context-independent alternatives conditions behaved like children in previous studies of scalar implicature: when all three animals were sleeping and children were asked, “Are some of the animals sleeping?” they did not strengthen the utterance. Children in this condition accepted 3-Item Test trials on 66.6% of trials. Their responses did not differ significantly between 2-Item True trials and 3-Item Test trials (t(14) = 1.0, p > .3), suggesting that they were equally likely to agree that some animals were sleeping when all three of them were as when only two animals were sleeping. The insertion of only did not...
on 80.0% of trials, and correctly denied that all of the `Are some of the animals sleeping?” (2-Item True trials)

natives (see sentences involving context-independent alternatives (children interpreted sentences involving contextual alter-
sized by the experimenter’s prosody.

First, consider the data for context-independent alter-

Fig. 2. Percentage of children who said “yes” to questions in the context-independent alternatives conditions. Error bars represent standard error.

Neither the word all nor the quantifier was empha-

The use of the word some, had a significant effect on how children’s responses in the context-independent alternatives conditions. Error bars represent standard error.

Fig. 3. Percentage of children who said “yes” to questions in the contextual alternatives conditions. Error bars represent standard error.

Some/Only some?          All?

Fig. 1. Some/Only some?          All?

Table 1. Did children accept 3-Item Test trials on 66.6% of trials. Their acceptance of 3-Item Test trials did not strengthen the utterance. Children in this condition were asked, “Are some of the animals sleeping?” they did not answer correctly: they agreed to answers on all three animals.

Barner, Brooks, & Bale (2011)

Context-independent alternatives

Contextually-defined alternatives

The results from this study are consistent with the hypothesis that children’s knowledge of scalar alternatives had a very large impact. As long as alternatives were provided contextually. For context-independent alternatives, no pragmatic inference was required. Thus, children’s failure to derive scalar implicatures in the experiment, children were able to accept the 3-Item test sentence if the sentence was grammatically strengthened and had contextual alternatives than if it was strengthened and added: “Are only the cat and the cow sleeping?”

Inference was required. Thus, children’s failure to derive scalar implicatures most of the time, even when they encountered when multiple dimensions of contrast are present in a context (see studies by Crain et al., in the Introduction).

Some/Only some?          All?         Some/Only some?

Cat & cow/Only cat & cow?

Cat & cow & dog?

Cat & cow/Only cat & cow?
Pragmatics: A developmental puzzle

• Children are experts at pragmatic word learning, but terrible at scalar implicature

• Pragmatic word learning requires intention reading (Bloom, 2002), knowledge of discourse status (Akhtar et al., 1996), and some type of perspective taking (Baron-Cohen, Baldwin, & Crowson, 1997)

• All of these are Grice-relevant abilities; what’s missing?
Searching for dissociations

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Underspecification</td>
<td>Said</td>
<td>Said</td>
<td>Explicature</td>
<td>Impliciture</td>
<td>Presumptive</td>
</tr>
<tr>
<td>Numeral upper-bounds</td>
<td>GCI</td>
<td>Said</td>
<td>Explicature</td>
<td>Impliciture</td>
<td>Presumptive</td>
</tr>
<tr>
<td>Ordering with <em>and</em></td>
<td>GCI</td>
<td>CI</td>
<td>Explicature</td>
<td>Impliciture</td>
<td>Presumptive</td>
</tr>
<tr>
<td>Lexical scalar inference</td>
<td>GCI</td>
<td>CI</td>
<td>Explicature</td>
<td>Impliciture</td>
<td>Presumptive</td>
</tr>
<tr>
<td>Particularized inferences</td>
<td>CI</td>
<td>CI</td>
<td>CI</td>
<td>CI</td>
<td>CI</td>
</tr>
</tbody>
</table>

- The project of distinguishing these theoretical classes is finding phenomena that dissociate them
- **What distinctions can we make?**