4.1.3 Results

For each pair, we predicted children would respond in one of three possible ways to the accompanying request: they would accept the request by giving one object, reject the request and give neither object, or reject the request and give both objects. Each of these three types of responses could potentially be correct depending on the pair, and predictions about each pair were clear, given the above discussion. We therefore chose to capture the results in terms of percentage of correct responses. This method is in contrast to capturing the results in terms of the percentage of responses in which the participant gave only one object (since such a percentage would not indicate which object was given) or the percentage of responses in which the participant only gave the object with the greater degree of the property (since in this case and the former, the percentage would not indicate what happened when the participant did not give one object, but gave both or neither). The results are presented in Table 6.

Table 6: Percentage of correct responses in Experiment 1a

<table>
<thead>
<tr>
<th>age</th>
<th>control items</th>
<th>test items</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>color</td>
<td>shape</td>
<td>mood</td>
<td>big</td>
<td>long</td>
<td>spotted</td>
</tr>
<tr>
<td>3 yrs</td>
<td>93</td>
<td>95</td>
<td>93</td>
<td>95</td>
<td>90</td>
<td>90</td>
</tr>
<tr>
<td>4 yrs</td>
<td>98</td>
<td>100</td>
<td>100</td>
<td>95</td>
<td>95</td>
<td>85</td>
</tr>
<tr>
<td>5 yrs</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>95</td>
</tr>
<tr>
<td>adults</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>98</td>
<td>98</td>
<td>98</td>
</tr>
</tbody>
</table>

These results show that both adults and children accepted and rejected the requests accordingly: both age groups were at ceiling with the control items and responded as predicted with the test items.
Recall that the pairs for the target stimuli also appeared as part of the series in the SJT. Children consistently judged items #1 and #3 in the relative sets (the bigger/longer pair in this task) as *big* or *long*, and they rejected this label for items #5 and #7 in the same series (the smaller/shorter pair in this task). In spite of these judgments, participants in the PAT did distinguish between the members of these pairs, treating *big*/*long* as true of #1 and false of #3 when confronted with this pair, and true of #5 and false of #7 when confronted with this pair. We take this as an indication that children were willing and able to shift the standard of comparison for the relative GAs. That they did not do so for the other items indicates that context-dependence is part of the lexical entry for these adjectives only and not the others.

The one instance in which children’s responses did not align with those of the adults was with the items corresponding to the maximum standard absolute GA *full*. Recall that there was a pair in which one of the containers was full and the other filled only to some degree. (We will call this pair the ‘full/non-full’ pair, abbreviated in the table above as ‘f/nf’). For another pair, there were two containers, neither of which was full, but one was fuller than the other. (We will call this pair the ‘non-full/non-full’ pair, abbreviated in the table above as ‘nf/nf’). Children’s correct responses to the ‘non-full/non-full’ pair are significantly lower than to the other target pairs (two-tailed t-tests t(29), ‘non-full/non-full’ v. *big* #1/3: t = -7.62, p < 0.00001; *big* #5/7: t = -5.29, p = 0.00001; *long* #1/3: t = -5.46, p < 0.00001; *long* #5/7: t = -7.62, p < 0.00001; *spotted* #1/4: t = -4.07, p = 0.0003; *spotted* #5/7: t = -7.62, p < 0.00001; ‘full/non-full’: t = -7.62, p < 0.00001). When presented with the ‘full/non-full’ pair, children consistently gave the puppet the full container; however, when presented with the ‘non-full/non-full’ pair, many children did not reject the request. Instead, they gave the puppet the fuller of the two containers.

At first glance, this type of response could be taken as indication that children are, in fact, interpreting the adjective in *the A one* as a comparative, which would call into question
our conclusions from the data involving relative GAs. However, children’s responses to the spotted objects argue against this conclusion. Given two spotted objects with unequal numbers of spots (e.g., items #1 and 4), children consistently rejected the puppet’s request for the spotted one, commenting that both objects were spotted, and did not give the puppet the more spotted member of the pair. The sharp difference between *full* and *spotted* suggests that whatever is going on here has to do with the way children understand *full*, rather than a general strategy for reanalyzing unmarked adjectives as comparatives in contexts that would otherwise result in presupposition failure.

One way in which children’s responses to the ‘non-full/non-full’ pair are curious is that the Scalar Judgment Task discussed earlier clearly showed that children share with adults a maximum standard absolute interpretation of the adjective *full*. In particular, 18 children who participated in the Presupposition Assessment Task also participated in the SJT within an interval of approximately three weeks. Judgments of the *full* SJT set from these children are presented in Table 7.\(^7\)

---

\(^7\) The analyses reported in this section compare responses in the SJT and the PAT just for the child participants who participated in both tasks in order to elucidate how performance in the latter task reveals interesting pragmatically-based reasoning in light of children’s semantic knowledge. A series of independent two-tailed t tests (df=28) comparing responses to the target stimuli in Experiment 1a for those children who did participate in the pre-experiment SJT and those who did not revealed either no difference whatsoever or a non-significant difference between the two groups, with the probability ranging from 0.45 to 0.77.
Table 7: Judgments of *full* from children participating in the SJT and the PAT

<table>
<thead>
<tr>
<th>response in SJT</th>
<th>children giving each response</th>
<th>age</th>
<th>3 yrs</th>
<th>4 yrs</th>
<th>5 yrs</th>
</tr>
</thead>
<tbody>
<tr>
<td>only #1 is full</td>
<td>12 (66.7%)</td>
<td></td>
<td>5</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>#1 and 2 are full</td>
<td>4 (22.2%)</td>
<td></td>
<td></td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>#1, 2 and 3 are full</td>
<td>2 (11.1%)</td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Recall that the ‘non-full/non-full’ pair was composed of items #4 and 6 from the set of seven items. *None* of these 18 children judged item #4 to be full in the SJT; however, in the PAT, 11 of these 18 children (61.1%) gave this item to the puppet to satisfy his request for *the full one*. (See Table 8.)

Table 8: Children in the SJT and the PAT who gave the puppet item #4

<table>
<thead>
<tr>
<th>response in SJT</th>
<th>children giving #4 as the <em>full one</em></th>
<th>age</th>
<th>3 yrs</th>
<th>4 yrs</th>
<th>5 yrs</th>
</tr>
</thead>
<tbody>
<tr>
<td>only #1 is full</td>
<td>8 (72.7%)</td>
<td></td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>#1 and 2 are full</td>
<td>2 (18.2%)</td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>#1, 2 and 3 are full</td>
<td>1 (9.1%)</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Given this pattern, we decided to probe children’s responses to the ‘non-full/non-full’ pair further, by looking to see whether the ordering condition to which participants were assigned had an effect. Indeed, upon further examination, we found that eight of the 11 children (72.7%) who gave the puppet the fuller container while having judged it to be *not full* in the SJT were in the condition in which this ‘non-full/non-full’ pair appeared early in the sequence. Moreover, every single one of the 15 children from Experiment 1a who were in this ordering condition gave the puppet the fuller of the two ‘non-full’ containers in response
to his request for the full one. Only five of the 15 children in the condition in which this pair appeared later in the sequence (after the ‘full/non-full’ pair) responded to the puppet’s request in this way (a difference of 100% and 33.3% between the two conditions).

This pattern leaves open two possibilities. The first is that children benefit from early exposure to an object that exemplifies the maximal degree of a property named by a maximum standard absolute GA in a task like this (e.g., seeing a full container). The second is that early exposure to relative adjectives in the test sequence creates a kind of priming effect, causing children to incorrectly treat full as relative rather than as maximum standard absolute. Whenever the ‘non-full/non-full’ pair appeared before the ‘full/non-full’ pair, it was also immediately preceded by a long pair; it is possible that in shifting the standard of comparison with the long pair to accommodate the presuppositions of the definite NP, children were then influenced to respond in a similar way to the ‘non-full/non-full’ pair that followed. This possibility is explored in Experiment 1b.

4.1.4 Discussion

The results of Experiment 1a demonstrate that both adults and children distinguish between three subclasses of gradable adjectives. They shift the standard of comparison for relative GAs, but do not for absolute GAs. Relative GAs such as big and long have as part of their semantic representation a contextually-determined standard of comparison, and so allow for the standard to shift from context to context. Accordingly, when presented with requests for the big/long one, participants willingly gave the bigger or longer member of the pair. In contrast, when presented with two spotted objects, participants did not give the more spotted object; this is to be expected, since spotted is a minimum standard absolute GA with a meaning along the lines of ‘have spots’. Adult participants similarly rejected requests for the full one when given two ‘non-full’ containers, a predicted response, since full is a maximum standard absolute GA. However, some children gave the puppet the fuller container in this
situation. This pattern contrasted with our findings from the SJT, in which we found that children did not judge the same container to be full. One possible explanation for this pattern of results is that a relative pair appearing earlier in the sequence affected children’s responses to the ‘non-full/non-full’ pair. This possibility is explored in Experiment 1b.

4.2 Experiment 1b

4.2.1 Introduction

The goal of Experiment 1b was to determine the source of the ordering effect observed in Experiment 1a. Specifically, we wanted to know if the prior presentation of a relative GA pair in the sequence of items influenced children to erroneously treat full as context-dependent, causing them to give the puppet the fuller of the two ‘non-full’ containers in response to his request for the full one.

4.2.2 Method

Participants

Seventeen children representing three age groups also participated in this task: six three-year-olds (3 boys 3 girls, range: 3;1 to 3;11, M: 3;5); six four-year-olds (2 boys 4 girls, range: 4;2 to 4;11, M: 4;6); and five five-year-olds (1 boy 4 girls, range: 5;2 to 5;10, M: 5;4). In addition, 10 adult native speakers of English (Northwestern undergraduates fulfilling an experimental requirement for a Linguistics course) served as controls.

Materials

The same objects from Experiment 1a were used. The only difference was in the sequence of items. In Experiment 1a, the ‘non-full/non-full’ pair was almost immediately preceded by a long pair, with only one control pair intervening. To evaluate the influence of the relative GA pair, we made a minor change in the order of presentation, simply switching the order of the long and ‘non-full/non-full’ pair so that the latter was no longer preceded by the former.
Meaning and Context in Gradable Adjectives

Procedure

The procedure was the same as in Experiment 1a.

4.2.3 Results

The results for Experiment 1b are presented in Table 9. The pattern of responses is similar to the one observed in Experiment 1a\(^8\): children and adults responded as predicted to the training, control, and test items, with the exception of the ‘non-full/non-full’ pair.

Table 9: Percentage of correct responses in Experiment 1b

<table>
<thead>
<tr>
<th>age</th>
<th>control items</th>
<th>test items</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>color</td>
<td>shape</td>
<td>mood</td>
<td>big</td>
<td>long</td>
<td>spotted</td>
<td>full</td>
</tr>
<tr>
<td>3 yrs</td>
<td>83</td>
<td>100</td>
<td>83</td>
<td>100</td>
<td>83</td>
<td>67</td>
<td>50</td>
</tr>
<tr>
<td>4 yrs</td>
<td>96</td>
<td>100</td>
<td>100</td>
<td>92</td>
<td>83</td>
<td>92</td>
<td>58</td>
</tr>
<tr>
<td>5 yrs</td>
<td>95</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>90</td>
<td>90</td>
<td>70</td>
</tr>
<tr>
<td>adults</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>95</td>
<td>85</td>
</tr>
</tbody>
</table>

Recall that the purpose of conducting Experiment 1b was to identify the source of children’s non-adult-like responses to the ‘non-full/non-full’ pair. A direct comparison between participants’ responses to this pair in both conditions of Experiment 1 is informative. As we

\(^8\) The lower rate of correct responses to the *spotted* stimuli by the three-year-olds may be explained by the fact that in Experiment 1b, the experimenter encouraged the child to give the puppet verbal feedback, but, unlike in Experiment 1a, did not encourage the child to give the puppet both objects when both fit the description. This change was not intentional.
observe in Table 10, both children and adults were at their best when this pair appeared later in the sequence. \(^9\)

Table 10: Percentage of correct responses to ‘non-full/non-full’ pair

<table>
<thead>
<tr>
<th>Age</th>
<th>Pair early, after relative pair (Experiment 1a)</th>
<th>Pair early, before relative pair (Experiment 1b)</th>
<th>Pair later, after f/nf pair (Experiments 1a, 1b)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>18</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>70</td>
<td>100</td>
</tr>
</tbody>
</table>

It appears that the prior presentation of a maximum standard (i.e., seeing the ‘full/non-full’ pair earlier in the sequence) allowed participants to be more successful in their judgments of the ‘non-full/non-full’ pair. Children were not misled by the prior presentation of a relative GA pair to treat the fuller container as the *full* one. These conclusions are supported by the fact that while the difference between either of the two ‘early’ orders and the ‘later’ order is significant, the difference between the two ‘early’ orders is not (Fisher’s Exact Tests, two-tailed probabilities: ‘early 1a’ v. ‘later’: p < 0.001; ‘early 1b’ v. ‘later’: p = 0.01; ‘early 1a’ v. ‘early 1b’: p = 0.24). However, knowing that children benefited from exposure to the maximum standard does not fully explain the results from Experiment 1. Even when given this boost, children’s responses to the ‘non-full/non-full’ pair still deviated from their adult-like responses to the other pairs. In order to explore this difference further we therefore turned to an additional analysis, in which we evaluated children’s reaction times (RTs) during the experimental session.

\(^9\) Adults who gave the fuller of the two ‘non-full’ containers noted at the end of the experimental session without any prompting that they realized this mistake later in the experiment and wished to make clear to the experimenter that they knew what *full* means.
The purpose of analyzing the RTs for children’s experimental sessions was to determine if children who gave the puppet the fuller of the two ‘non-full’ containers in response to his request for the full one took longer to do so for this pair than with other pairs. We reasoned that children might have evaluated the puppet’s request given the experimental stimuli at hand and subsequently decided to treat it as being an instance of IMPRECISION: informative use of a false description to identify the object that comes closest – and close enough – to satisfying it. That is, although children might not have actually judged the fuller of the two ‘non-full’ containers to be full (maintaining an absolute maximum semantics for the adjective along the lines of ‘be maximally full’), they still might have decided after some deliberation that they should tolerate some ‘loose talk’ on the part of the puppet and hand over the container that came closest to satisfying the description that he used in his request. Crucially, this kind of reasoning is not necessary in the case of relative GAs (such as big and long), since the presuppositions of the definite description can be accommodated, and the request therefore accepted, based entirely on the (context sensitive) semantics of the adjective. This kind of reasoning is not applicable in the case of minimum standard GAs (e.g., spotted), which apply to any object that has the relevant property to some degree, and so are inherently imprecise. It follows that if this reasoning is brought into play just for the purpose of deciding how to respond to a pair of objects (neither of which satisfies the semantic requirements of a maximum standard absolute GA), we should see an effect on RTs just in such cases.

Experimental sessions with child participants were videotaped using a Sony Digital8 Handycam. Videotapes were imported from the camera into Apple Inc’s iMovie program as .mov files. Videos were then coded offline by research assistants in our laboratory on a
Macintosh computer using SuperCoder software (Hollich, 2003). For each item in which the child accepted the puppet’s request (i.e., gave the puppet one of the two objects), the research assistants coded three measurements: the child’s look to the object, the child’s reach toward the object, and the child’s touch of the object. These RT measurements were then coded frame by frame, where one frame is equal to 1/30 of a second.

It is necessary to provide two additional details about the coding session. First, we excluded from analysis any items in which the children’s eye movements could not be coded (e.g., if the eyes were occluded), the child was already looking at or touching the stimuli before the request was uttered, or any other experimental artifact prevented the coders from obtaining measurements (i.e., there was a distraction in the background). For this reason, the total number of children whose RTs were analyzed varies from analysis to analysis. This number is always provided in a footnote. Second, rather than coding the initial look to the object, since the child could have decided to inspect the second object before deciding to give the puppet the first object, we coded the look that immediately preceded the reach to the object. A reach was a movement that ultimately resulted in touching an object. We chose to target these measurements instead of, e.g., proportion of looking time, since we wished to measure latency of response across key items.

We then targeted two sets of RT measurements for analysis. In the first, we asked if children took longer to respond to the puppet’s request when it accompanied the ‘non-

---

10 At least two coders were assigned to each experimental session, with one coder arbitrarily chosen as the default. The inter-coder rate of agreement across all trials averaged above 95%. In case there was a disagreement of more than 5 frames for any of the three measurements, a third coder was brought in as a tiebreaker for that item.
full/non-full’ pair than when it accompanied the ‘full/non-full’ pair. These results are presented in Figure 3.\textsuperscript{11}

Figure 3: Reaction times for two full pairs in Experiment 1a

\textsuperscript{11} We analyzed the RTs for 16 children who gave the puppet the fuller container for each pair.

Indeed, differences between the look, reach, and touch are significant for these two pairs (one-tailed t tests: look $t(15) = 1.71, p = 0.05$; reach $t(15) = 3.03, p = 0.004$; touch $t(15) = 3.47, p < 0.002$).

In the second RT measurement, we compared the RTs for the ‘non-full/non-full’ pair to those for other key pairs in order to determine if children generally took extra time to shift the standard of comparison, or if the extra time observed with the ‘non-full/non-full’ pair was unique to that pair. We targeted the difference between the look and the touch of the object
for four key pairs, the two big pairs (two big blocks and two small blocks) and the two full pairs. These results are presented in Figure 4.\textsuperscript{12}

Figure 4: Reaction times for four key pairs in Experiment 1a

The RTs for the ‘non-full/non-full’ pair clearly stand out from the three others. Indeed, while the RT for this pair is significantly longer than every other pair (two-tailed t-tests: ‘non-full/non-full’ v. ‘full/non-full’: t(48) = 2.42, p = 0.02; ‘non-full/non-full’ v. ‘big’: t(45) = 3.07, p < 0.01; ‘non-full/non-full’ v. ‘small’: t(51) = 3.79, p < 0.001), the other pairs do not differ significantly from each other (two-tailed t-tests: ‘full/non-full’ v. big: t(53) = 1.11, p = 0.27; ‘big’ v. ‘small’: t(56) = 0.87, p = 0.39; ‘full/non-full’ v. ‘small’: t(59) = 1.96, p = 0.05, marginally significant). These results demonstrate that the shift in the standard of comparison for the big pairs – even for the pair in which the small blocks were judged to be

\textsuperscript{12} We analyzed all codable responses for these items across all children. The number of children varied for each item: 26 for the big blocks, 32 for the small blocks, 29 for the ‘full/non-full’ pair, and 21 for the ‘non-full/non-full’ pair.
not big in the SJT – was automatic. This is expected: given that a contextually-determined standard is part of the semantic representation for big, no additional time should be required to compute the standard of comparison.

4.2.4 Discussion

The results of Experiment 1b elaborate upon those of Experiment 1a by demonstrating that the source of children’s non-adult-like performance with the ‘non-full/non-full’ pair (i.e., giving the puppet the fuller container in response to his request for the full one) is not driven by the influence of a preceding relative GA pair. Children are significantly more likely to pattern like adults with the ‘non-full/non-full’ pair if it occurs later in the sequence, some time after they have seen the maximum standard exemplified, but even with this assistance, some children are still inclined to allow the fuller container to count as the full one. The RT analysis shed light on this pattern of response by demonstrating that children take significantly longer to give the puppet the fuller of the two ‘non-full’ containers, suggesting that additional reasoning beyond that involved in assigning context (in)dependent meanings to GAs is, in fact, being brought to bear on these examples. We will provide a more detailed analysis of the significance of the RT data in the General Discussion section.

5. Experiment 2

5.1 Experiment 2a

5.1.1 Introduction

Given children’s responses to the full pairs in Experiment 1, one question immediately surfaces: are these responses unique to full, or can they be generalized to a larger

---

13 One might have expected a semantic congruity effect with the small pair of blocks (i.e., longer RTs resulting from the fact that the adjective in the request was big, but the blocks were judged to be not big). The lack of such an effect here may be due to the fact that the difference in magnitude between the two pairs was not extraordinarily large.
set of maximum standard absolute GAs? The goal of Experiment 2 was to address this question.

5.1.2 Method

Participants

Thirty children representing three age groups participated in this task: 10 three-year-olds (5 boys 5 girls, range: 3;2 to 3;11, M: 3;6); 10 four-year-olds (5 boys 5 girls, range: 4;1 to 4;10, M: 4;4); and 10 five-year-olds (4 boys 6 girls, range: 5;0 to 5;7, M: 5;3). In addition, 24 adult native speakers of English (Northwestern undergraduates fulfilling an experimental requirement for a Linguistics course) served as controls.

Materials

The materials were the same as in Experiments 1, with the exception of four target pairs. In place of the two *full* pairs, there were two pairs corresponding to another maximum standard absolute GA, *straight*. In place of the two *spotted* pairs, there were two pairs corresponding to another minimum standard absolute GA, *bumpy*. These pairs were designed similarly to those in Experiment 1, so that only one of the two pairs for each adjective would satisfy the presuppositions of the definite description. Details for these stimuli are presented in Table 11.

---

The fact that *full* and *straight* are both absolute maximum standard absolute GAs can be highlighted by adverbial modification. While both of these adjectives may be modified by either *completely* or *perfectly*, the relative GAs *big* and *long* cannot. Likewise, the fact that *spotted* and *bumpy* are both minimum standard absolute GAs can be highlighted by modifying them by *slightly*, an adverb that cannot appear with the other target GAs presented at test.
Table 11: Stimuli substituted for the absolute GA pairs from Experiment 1

<table>
<thead>
<tr>
<th>adjective</th>
<th>pragmatic status of request</th>
<th>stimuli</th>
</tr>
</thead>
<tbody>
<tr>
<td>straight</td>
<td>felicitous</td>
<td>2 wire rods, 9” in length, one completely straight and the other bent into a “C”</td>
</tr>
<tr>
<td>(in place of full)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>infelicitous</td>
<td></td>
<td>2 wire rods, 9” in length, one curly and the other even curlier</td>
</tr>
<tr>
<td>bumpy</td>
<td>felicitous</td>
<td>two flat wooden boards, measuring 2” x 5”</td>
</tr>
<tr>
<td>(in place of spotted)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>infelicitous</td>
<td></td>
<td>two flat wooden boards, measuring 2” x 5”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>and painted orange, one with many bumps and one with less bumps</td>
</tr>
</tbody>
</table>

Procedure

The procedure was the same as in Experiment 1.

5.1.3 Results

As in Experiment 1, children and adults patterned as predicted for all items, with the exception of the maximum standard absolute GA stimuli. The same difference observed in Experiment 1 is found between the two straight pairs (abbreviated in the table as ‘b/b’ for
‘bent/bent’ and ‘s/b’ for ‘straight/bent’). The percentages of correct results are presented in Table 12.

Table 12: Percentage of correct responses in Experiment 2a

<table>
<thead>
<tr>
<th>age</th>
<th>control items</th>
<th>test items</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>color</td>
<td>shape</td>
</tr>
<tr>
<td>3 yrs</td>
<td>83</td>
<td>90</td>
</tr>
<tr>
<td>4 yrs</td>
<td>95</td>
<td>100</td>
</tr>
<tr>
<td>5 yrs</td>
<td>95</td>
<td>100</td>
</tr>
<tr>
<td>adults</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

The same ordering effect observed in Experiment 1 was also found. Of the seven children across age groups who gave the straighter rod for the ‘bent/bent’ pair, six (or 85.7%) of them saw this pair early in the sequence, well before they were exposed to the ‘straight/bent’ pair.

Now, while the percentage of correct responses to the ‘bent/bent’ pair was in the same direction as the ‘non-full/non-full’ pair, the percentage of correct responses to this pair was not nearly as low as in Experiment 1. In Experiment 1a, only 33% of the children patterned as predicted with the ‘non-full/non-full’ pair by rejecting the request, and in Experiment 1b, only 19% of the children did so. By contrast, in this experiment, 77% of the children rejected the request. However, participants’ responses to this pair are still significantly lower than to most (but not all) of the other test pairs (two-tailed t-tests, t(29): ‘bent/bent’ v. big #1/3: t = -2.97, p = 0.006; big #5/7: t = -2.26, p = 0.03; long #1/3: t = -2.97, p = 0.006; bumpy #5/7: t = -2.97, p = 0.006; ‘straight/bent’: t = -2.97, p = 0.006; but v. bumpy #1/4: t = 1.99, p < 0.06, marginally significant; and v. long #5/7: t = -1.28, p = 0.211, n.s.).
5.1.4 Discussion

The results of Experiment 2a support a generalization of the results from Experiment 1b to a larger set of absolute GAs, since child and adult responses to *straight* and *bumpy* were similar to their responses for *full* and *spotted*, respectively. Both age groups rejected the infelicitous requests for the items corresponding to these GAs and did not shift standards as they did for the relative GA pairs in Experiments 1 and 2. We observed, though, that children were more likely to reject the infelicitous request for the ‘bent/bent’ pair than they were for the ‘non-full/non-full’ pair. However, it is possible that the higher percentage of correct responses for the ‘bent/bent’ pair was an effect of the stimulus design. We observed *post facto* that the objects for *straight* shown in Table 11 seem to correspond better to stimuli for a minimum standard absolute GA such as *bent* or *curly* than for the maximum standard absolute GA *straight*: while item #1 was completely straight, none of the other wires had any degree of straightness. This design differed from the *full* materials, where it was the case that while only container #1 was full, the others in the PAT manifested some degree of being filled. This observation led us to conduct Experiment 2b using a minimally revised set of stimuli.

5.2 Experiment 2b

5.2.1 Introduction

The goal of this experiment was to introduce a minimally revised set of *straight* materials that more closely resembled the design of the *full* materials in Experiment 1, in order to determine whether participants’ judgments about the *straight one* when shown the ‘bent/bent’ pair would resemble those in Experiment 1a and 1b more than those in Experiment 2a did.
5.2.2 Method

Participants

Thirty children representing three age groups participated in this task: 10 three-year-olds (3 boys 7 girls, range: 3;2 to 3;10, M: 3;6); 10 four-year-olds (4 boys 6 girls, range: 4;1 to 4;9, M: 4;7); and 10 five-year-olds (5 boys 5 girls, range: 5;1 to 5;11, M: 5;6). In addition, 24 adult native speakers of English (Northwestern undergraduates fulfilling an experimental requirement for a Linguistics course) served as controls.

Materials

The materials were the same as in Experiment 2a, with the exception of the *straight* pairs. In place of the three wire rods that were bent to varying degrees, we inserted rods that resembled more accurately the containers used for *full*. For example, a completely straight rod was paired with a rod that was straight for most of its length but which had a curl at the top, analogous to a container which is filled most, but not all, of the way. The new pairs are presented in Table 13.

Table 13: Stimuli substituted for the *straight* pairs from Experiment 2a

<table>
<thead>
<tr>
<th>adjective</th>
<th>pragmatic status of request</th>
<th>stimuli</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>straight</em></td>
<td>felicitous</td>
<td>a completely straight rod and a mostly-straight rod with a curly section at the top</td>
</tr>
<tr>
<td></td>
<td>infelicitous</td>
<td>a rod with a long straight section and a curly section at the top and a totally curly rod</td>
</tr>
</tbody>
</table>
Procedure

The procedure was the same as in the previous experiments.

5.2.3 Results

In Experiment 2b, we observe the same trend as in each of the previous PAT experiments, with children and adults responding as predicted for all of the training, control, and test items, with the only exception being the ‘bent/bent’ stimuli. These results are presented in Table 14.

Table 14: Percentage of correct responses in Experiment 2b

<table>
<thead>
<tr>
<th>age</th>
<th>control items</th>
<th>test items</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>color</td>
<td>shape</td>
</tr>
<tr>
<td>3 yrs</td>
<td>88</td>
<td>80</td>
</tr>
<tr>
<td>4 yrs</td>
<td>90</td>
<td>90</td>
</tr>
<tr>
<td>5 yrs</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>adults</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

This time, children’s responses to the ‘bent/bent’ pair are more like those seen in Experiment 1 than those in Experiment 2a were and are significantly lower than to all of the other test pairs (two-tailed t-tests, t(29): ‘bent/bent’ v. big #1/3: t = -5.04, p < 0.0001; big #5/7: t = -4.47, p = 0.0001; long #1/3: t = -5.39, p < 0.00001; long #5/7: t = -3.27, p < 0.01; bumpy #1/4: t = -1.88, p < 0.07; bumpy #5/7: t = -4.10, p < 0.001; ‘straight/bent’: t = -3.53, p = 0.001).

In this experiment, 53% of the children rejected the request for the straight one when presented with the ‘bent/bent’ pair (compared to the 77% with the previous stimuli in Experiment 2a and 33% for the similarly-designed ‘non-full/non-full’ pair with the same orders of presentation in Experiment 1a). Again, there was also an effect of order of
Meaning and Context in Gradable Adjectives

presentation: children were more likely to reject the puppet’s request if they saw the ‘bent/bent’ pair later in the sequence; nine of the 13 children (69%) who gave the puppet the straighter of the two bent rods saw this pair appeared early in the sequence of items. The comparison presented in Figure 5 highlights the consistency of this ordering effect across Experiments 1 and 2.

Figure 5: Rate of acceptance of the non-maximal member in Experiments 1 and 2

This between-experiment comparison allows us to make the following generalization: when the pair with the object exemplifying the maximum degree appears later in the sequence than the pair with both non-maximal objects, children are significantly more likely to incorrectly accept the puppet’s request for the maximal object when shown the earlier pair. A series of single-factor ANOVAs for each experiment supports this conclusion (Experiment 1: F(2, 44) = 3.21, p < 0.0001; Experiment 2a: F(1, 28) = 4.20, p = 0.03; Experiment 2b: F(1, 28) = 4.20, p = 0.07, marginally significant). (See also the statistical analysis following the data presented in Table 10 for differences between the two early orders and the late order in Experiment 1.)
Recall that in Experiment 1, we conducted an analysis of children’s RTs as they responded to the puppet’s request. This analysis provided us with evidence that children took longer to accept the request (i.e., give the puppet the fuller container) when shown the ‘non-full/non-full’ pair than when shown either the ‘full/non-full’ pair or both big pairs. Although there were only five children for which the comparison between the RTs for the ‘straight/bent’ and ‘bent/bent’ pairs could be made, we observed the same trend in RTs: the differences between the adjective onset to the look, reach, and touch between the two pairs were all significant (one-tailed t tests: look t(4) = -3.23, p = 0.016; reach t(4) = -3.15, p = 0.017; touch t(4) = -3.32, p = 0.015). See Figure 6.

Figure 6: Reaction times for two *straight* pairs in Experiment 2b

---

15 We targeted the responses from Experiment 2b only and did not combine responses across both experiments, as we did in Experiment 1, since the *straight* stimuli differed between the two experiments. Recall, too, that we were constrained with respect to which videos we could code, for reasons outlined earlier. The total number of codable sessions and items was therefore less than for Experiment 1.
As in Experiment 1, we also compared RTs for the difference between the look and the touch for the ‘bent/bent’ pair to three other key pairs – the ‘straight/bent’ pair and the two pairs of blocks corresponding to the request for the big one. We again observed the same trend as in Experiment 1 (cf. Figure 7): children took significantly longer to select the straighter of the two bent pairs to satisfy the puppet’s request than they did for the other three pairs (two-tailed t-tests: ‘bent/bent’ v. ‘straight/bent: t(23) = 2.15, p = 0.04; ‘bent/bent’ v. ‘big’: t(20) = 2.31, p = 0.03; ‘bent/bent’ v. ‘small’: t(22) = 2.75, p = 0.01). And, as before, the other pairs do not differ significantly from each other (two-tailed t-tests: ‘straight/bent’ v. ‘big’: t(31) = -0.33, p = 0.74; ‘big’ v. ‘small’: t(30) = 0.64, p = 0.53; ‘straight/bent’ v. ‘small’: t(33) = 0.86, p = 0.40).16

Figure 7: Reaction times for four key pairs in Experiment 2b

5.2.4 Discussion

The results of Experiment 2 support a generalization of the results from Experiment 1. In particular, the results show that the effect of the order of presentation and the longer RTs

16 The data from the following number of children were analyzed for each pair: 15 for the big blocks, 17 for the small blocks, 18 for ‘straight/bent’, and 7 for ‘bent/bent’.
for the non-maximal pair are not unique to the stimuli included in Experiment 1, and therefore not unique to the lexical item full. Rather, the fact that we see the same pattern with the absolute maximum standard GA straight indicates that the interpretive processes involved in choosing an object that comes closest to satisfying a description based on a maximum standard GA are different from those involved in choosing an object that satisfies a context-dependent description based on a relative GA. In short, understanding the big/long one as a description of the bigger/longer of two objects of unequal size is automatic, while tolerating use of the full/straight one as a description of the fuller/straighter of two objects when neither of them is actually full or straight requires some work.

6. General Discussion

The experiments reported in this paper provide new insights into children’s competence in three different aspects of the context/meaning relation. First, our study shows that by three years of age, children have adult-like competence in assigning context-sensitive interpretations to relative GAs in the positive form. Specifically, they are able to shift the standard of comparison for such adjectives in a way that is appropriate for the context of utterance. In our case, this was determined by the pragmatic demands of the definite description in which the adjective appeared, thereby changing the extension of the predicate. Moreover, the fact that children did not assign the same kinds of context-sensitive meanings to absolute GAs in the same contexts shows that they have already made adult-like distinctions between predicates that have context-sensitive denotations and those that do not, but which are otherwise semantically quite similar. In the case at hand, both relative and absolute GAs share the fundamental feature of expressing a relation to a scalar concept – that
is, both are gradable – but they differ just on the basis of the context dependence of the positive form.\textsuperscript{17}

Second, our experiments show that children as young as three have adult-like competence in dealing with the existence and uniqueness presuppositions of a singular definite description. Their responses to the control items in the PAT clearly show that they are aware of these presuppositions, and that they are willing to reject as infelicitous utterances that violate them. For example, in rejecting the puppet’s request, children often offered statements such as the following: \textit{The very big one?}; \textit{Oh, but I have TWO red ones!}; \textit{What red one? He should say what shape!}; \textit{He thinks there must be two different colors!}; \textit{They’re all spotted!}; \textit{The very bumpy one?}; \textit{What bumpy one?}

At the same time, their responses to the test items containing relative GAs show that when the semantics of the adjective and the context of utterance allow for the possibility of presupposition accommodation – in these cases, by shifting the standard of comparison so that the adjective (and, consequently, the description) is true of just one member of a pair of objects – children, like adults, will accommodate. These two results together indicate that (at least in this domain) children are constructing the type of complex discourse models that linguistic expressions with presuppositions must be integrated into, and are willing and able to automatically update those models to allow integration when such a move is licensed.

\textsuperscript{17} According to Kennedy & McNally (2005) and Kennedy (2007), this difference actually stems from a more basic difference between the two classes of GAs, namely the structures of the scales that represent the gradable concepts they encode: relative GAs use open scales, while absolute GAs use closed scales. (See also Rotstein & Winter (2004).) If this is correct, then a more accurate way to state the generalization is that by three years of age, children already know the mapping between scale structure and context sensitivity of the positive form. See Syrett (2007) for extensive discussion of this issue.
Finally, our results provide important new data from child language that bears on fundamental questions about the nature of contextually influenced interpretive variability. The results from the PAT show that children will treat both relative GAs and maximum standard absolute GAs as ‘variable’: just as the big one was accepted as a description of the bigger of two blocks, even when it was otherwise judged to be not big, the full one was sometimes accepted as a description of the fuller of two containers, even when it was observably not full. The latter result was significantly less frequent, and was influenced by the sequence of requests, but it nevertheless occurred enough to support the conclusion that some sort of interpretive variability is at work. The question at hand is whether this variability is due to the same factors that are involved in variable interpretations of relative GAs – semantic context dependence (and ultimately perhaps the semantic features that give rise to vagueness) – or whether it stems from a different source.

It is important to emphasize that this is a much more general question, and bears on both child and adult language. While we did not see the same variability in maximum standard absolute GAs with adults, we suspect that we would have if the materials had been slightly different. In particular, we believe that if the fuller of the two ‘non-full’ containers had been closer to full than it actually was, though still noticeably not full, adults would have behaved like children and provided it in response to a request for the full one. Our belief is based on simple observation of everyday linguistic behavior: we regularly speak imprecisely. Thus, in addition to using absolute GAs like full and straight to describe objects that are not strictly speaking maximally full or maximally straight, we might also describe our arrival somewhere at 3:05 p.m. by saying We arrived at 3 p.m., or a child’s consumption of all but a few pieces of a sandwich by saying The child ate the sandwich. (See Lasersohn (1999) and Sperber & Wilson (1986) for extensive discussion of such cases.) Given that our willingness to tolerate such imprecision is itself a matter of context (i.e., if the precise time of our arrival
is important, then use of *arrive at 3pm* to describe an arrival at 3:05 p.m. is inappropriate), it could be the case that imprecision is just another instance of semantic context dependence. In other words, it could be the case that interpretive variability is always fundamentally semantic, and that expressions like *full* and *straight* (as well as expressions such as *arrive at 3 p.m.*, *eat the sandwich*, etc.) have meanings that, like *big* and *long*, require valuing some contextual parameter before they can be assigned actual extensions. Such meanings might be slightly different for absolute maximum GAs – *full*, for example, might have a meaning along the lines of ‘have a fullness that is at least as close to maximally full as required by the prevailing contextual standard of precision’, but on this view, they would be of the same basic sort as the context sensitive meanings of relative GAs.

While this conclusion would be consistent with children’s gross responses in the PAT (whether they gave the puppet an object to satisfy his request or not), it is not consistent with the reaction time data that we collected. Instead, the fact that children systematically took significantly longer to respond to a request for *the full/straight one* in contexts involving two objects that did not satisfy the maximum standard property (that is, which were not maximally full or straight) shows that processing of more than just the semantic content of the description was involved in their responses. To see why this is the case, consider the fact that their response times to definite descriptions based on relative GAs (e.g., *the big/long one*) were the same regardless of whether they were looking at two objects from the upper end of the continuum (both objects judged *big/long*, based on the Scalar Judgment Task) or two objects from the lower end (neither object judged *big/long* in the SJT). This has a straightforward explanation. Focusing on the case of *long* for purposes of exemplification, if children know the meaning of the relative GA, then they know that it denotes the property of having a length that exceeds the standard of comparison for length in the context. If the context contains exactly two relevant objects and is such that just one of them has this
property (a constraint imposed by the presuppositions of the definite description used by the speaker), their response is straightforward: choose the longer object. It is unnecessary for them to choose some particular degree as the standard of comparison, because given two objects of unequal lengths, the transitivity of the ‘exceeds’ relation ensures that if just one of the objects has a length that exceeds the standard of comparison (whatever it may be), it has to be the longer of the two.

Crucially, if the same kind of reasoning were at play in children’s evaluation of requests for the full one in the ‘nonfull/nonfull’ context, we should have seen similar RTs. Specifically, if full meant ‘fuller than the contextual standard’ or even ‘at least as close to maximally full as the contextual standard’, then the choice between two objects of unequal fullness would again be straightforward: if context dictates that just one of the objects should satisfy the property, it will always be the fuller of the two. The fact that children took significantly longer to respond to such requests, even though their responses were the same (i.e., they chose the object with the greater degree of the relevant property), clearly shows that the same kind of reasoning is not at work in evaluating absolute GAs in such contexts, and therefore that the kind of interpretive variability that such adjectives do occasionally display is not indicative of a context-dependent semantics.

But what kind of reasoning do the long RTs indicate? Our hypothesis is that children took longer to respond to such cases precisely because they were aware that neither of the objects in question actually satisfied the description used in the request, and that this knowledge triggered an evaluation of whether one of the objects was actually close enough to having the property in order for them to treat the speaker’s description as though it were a description of that object. There are different ways to formally characterize this notion of imprecision. For concreteness, we will assume with Lasersohn (1999) that it involves (a) computation of a set of alternative denotations for an expression that are ordered relative to
the actual meaning, and (b) a decision about how far down this list one can go before getting too far away from the actual denotation. In contexts in which the actual meaning of an utterance fails to apply, if there is an element among the set of its ‘tolerable alternatives’ that could actually be true, the utterance can be taken to count as ‘close enough to true’ for the purpose of the discourse, and treated as though it had been an utterance of the tolerable alternative. Crucially, this kind of reasoning is fundamentally pragmatic in nature, as it involves judgments about communicative intent and whether a particular utterance can be used in a way that (strictly speaking) conflicts with the truth conditional requirements that it imposes by virtue of its semantics. It is this move beyond the computation of semantic content – even content that invokes contextual information – that, we claim, imposes an extra processing load and results in the longer RTs we observed.

If this is correct, then we have experimental evidence for a distinction between two types of context-dependent interpretive variability. One type – that exhibited by relative GAs in the positive form – is fundamentally semantic in nature, and is based on the conventional meaning of particular expressions (or combinations thereof). A second type – that exhibited by imprecise uses of maximum standard absolute GAs – is fundamentally pragmatic, and involves computation of a set of alternative meanings and a contextual judgment about how many of them count as ‘tolerable deviations’ from the actual, precise meaning of the expression. The differences between children and adults that we observed in their willingness to accept false descriptions based on maximum standard GAs can be explained by assuming that children are more willing to tolerate imprecision than adults, at least in this experimental task.\(^\text{18}\) If our overall interpretation of the data is correct, then we predict that smaller

\(^{18}\) One potential objection to this interpretation of the results is that preschoolers are generally not imprecise, and can in fact be quite literal at times. (Thanks to Anastasia Giannakidou, p.c., for raising this point.) Why, then, would they tolerate imprecision in the PAT? In fact,
deviations from the absolute standard should result in adults behaving like children – accepting false descriptions based on absolute maximal standard GAs, but taking longer to do so than the time it takes to accept a description based on a relative GA.

7. Conclusion

In sum, the research described above provides evidence for four kinds of conclusions. First, by the age of three, children have fine-grained distinctions among subclasses of gradable adjectives (GAs). In particular, we have seen that children distinguish relative GAs such as big from absolute GAs such as spotted/full. Evidence for this conclusion comes from children’s recognition that that standard of comparison against which a GA in the positive form is judged varies only for relative GAs such as big. Second, children distinguish maximam standard absolute GAs such as full from minimum standard absolute gradable adjectives such as spotted. Evidence for this conclusion comes from the Scalar Judgment Task in which objects exhibiting the maximal degree of fullness were judged as full but objects exhibiting any degree of spottedness were judged as spotted. Third, we learned that children have a complete understanding of both the uniqueness and existence presuppositions associated with singular definite descriptions in English. We would not have been able to find the context dependence of relative GAs were that not the case. It is only through these presuppositions that appropriate shifting of the standard of comparison is licensed. Fourth, we are not claiming that children are imprecise; indeed, the slower RTs for the ‘non-full/non-full’ and ‘bent/bent’ pairs indicate that their willingness to give the puppet the fuller or straighter member of the pair is far from automatic. Our claim is rather that their desire to maintain a high standard of precision was, in this case, overridden by a stronger desire to actually respond to the speaker’s (the puppet’s) request. Evidently, the opposite is true for adults, though as we suggested above, we suspect that the results would change if the deviation from the maximum were reduced.
have identified novel evidence for the distinction between context dependence as a function of the meaning of an expression as opposed to context dependence as a function of the use of an expression in a situation. We have argued that the reaction time difference between accepting the bigger of two small blocks as *big* and the fuller of two non-full containers as *full* is a direct reflection of the different computations involved in these two types of context dependence, the former being semantic in nature and the latter pragmatic.

These results open a range of questions for future research. With respect to the subcategories of adjective, it is important to determine what kinds of evidence lead learners to classify a given adjective as gradable, relative, or absolute. To what extent do these distinctions follow directly from the mapping of an adjective to an appropriate concept as opposed to being linguistically conditioned? Even if the subclasses of GAs represent limitations on the range of possible scalar meanings, what kinds of evidence do learners use to place a novel adjective in one of these subclasses? (See Syrett (2007) for an account of how infants can recruit at least one surface-level cue, that of adverbial modification, to initially classify novel adjectives into these different subclasses.) These questions would seem to interact with the issues surrounding context dependence. If both relative and absolute GAs show some degree of context dependence, though in different ways, what information does the learner use to determine whether this context dependence is due to semantics or pragmatics? Indeed, what evidence leads the learner to recognize the distinction and to attribute some cases of context dependence to the semantics and other cases to the pragmatics?

While these results have raised as many questions as they have answered, we believe they also provide an interesting methodological lesson. In particular, while the debate concerning varieties of context dependence has raged for years within linguistics and philosophy, resisting conclusive evidence, the current experiments allow us to ask about the
real time consequences of the alternative approaches. Using experimental evidence from children we are able to see the alternative computations involved in semantic vs. pragmatic context-dependence unfolding in real time. Our work demonstrates that the process of verifying of the meaning of an expression can provide evidence for alternative hypotheses about the meaning itself. Moreover, such evidence can come equally well from children as from adults.
Author’s addresses

Kristen Syrett
Rutgers Center for Cognitive Science (RuCCS)
Psychology Building Addition, Busch Campus
Rutgers University – New Brunswick
152 Frelinghuysen Rd.
Piscataway, NJ 08854-8020

Christopher Kennedy
Department of Linguistics
University of Chicago
1010 E. 59th St.
Chicago, IL 60637

Jeffrey Lidz
Department of Linguistics
1401 Marie Mount Hall
University of Maryland
College Park, MD 20742-7505
References


Meaning and Context in Gradable Adjectives


Klibanoff, R. S. & S. R. Waxman (1998), Preschoolers’ acquisition of novel adjectives and the role of basic-level kind. In A. Greenhill, M. Hughes, H. Littlefield and H. Walsh


Wexler, K. (2003), Determiners in development, why there is a problem: Maximal trouble. Handout from *Semantics in acquisition* (course taught at MIT).