DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD

Semantic Meaning and Pragmatic Interpretation in 5-Year-Olds: Evidence from Real-Time Spoken Language Comprehension

Citation	Huang, Yi Ting, and Jesse Snedeker. 2009. Semantic meaning and pragmatic interpretation in 5-year olds: Evidence from real-time spoken language comprehension. Developmental Psychology 45(6): 1723-1739.
Published Version	doi:10.1037/a0016704
Accessed	February 19, 2015 10:53:00 AM EST
Citable Link	http://nrs.harvard.edu/urn-3:HUL.InstRepos:9978129
Terms of Use	This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Open Access Policy Articles, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#OAP

(Article begins on next page)

Semantic meaning and pragmatic interpretation in five-year-olds: Evidence from real time spoken language comprehension

Yi Ting Huang[1] and Jesse Snedeker[2]
[1] Department of Psychology, The University of North Carolina at Chapel Hill
[2] Department of Psychology, Harvard University

Date resubmitted: May 05, 2009

Abstract

Recent research on children's inferencing has found that while adults typically adopt the pragmatic interpretation of *some* (implying *not all*), five- to nine-year-olds often prefer the semantic interpretation of the quantifier (meaning possibly *all*). Do these failures reflect a breakdown of pragmatic competence or the metalinguistic demands of prior tasks? In three experiments, we used the visual-world eye-tracking paradigm to elicit an implicit measure of adult's and children's abilities to generate scalar implicatures. While adults' eye-movements indicated that they had interpreted *some* with the pragmatic inference, children's looks suggest that they persistently interpreted *some* as compatible with *all* (Experiment 1). Nevertheless, both adults and children were able to quickly reject competitors that were inconsistent with the semantics of *some*, confirming the sensitivity of the paradigm (Experiment 2). Finally, adults, but not children, successfully distinguished between situations that violated the scalar implicature and ones that did not (Experiment 3). These data demonstrate that children interpret quantifiers based on their semantic content and fail to generate scalar implicatures during online language comprehension.

Keywords: language development, semantics, pragmatics, scalar implicature, quantifiers

1. Introduction

To become linguistically competent, children must be able to not only understand the literal content of utterances but also make appropriate inferences that capture the speaker's intended meaning. For example, in the dialogue in (1), we can infer from Mother's response that the Child is not allowed to eat the entire cake.

(1)   Child: Can I please eat the cake?

        Mother: You can have a slice.

Nevertheless, eating a slice of cake does not rule out the possibility of finishing it. In fact, one event typically precedes the other. Thus we can imagine a situation where our initial inference (I ate a slice but no more) is explicitly canceled by a subsequent statement in (2).

(2)   Mother: Did you eat a slice of the cake?

        Child: Yeah, I ate a slice. In fact, I ate the whole thing.

This division between a linguistically-encoded meaning and the inferences that we can derive from it was made prominent by Grice (1957/1975). *Semantics* refers to the aspects of the interpretation that can be directly calculated from the meanings of words and the structural relationships between them. In contrast, *pragmatics* refers to the aspects of interpretation that are inferred through an analysis of the context and the communicator's goals. Grice proposed that while pragmatic inferences make use of the semantic content of an utterance, they are distinct from truth conditional meaning since they are frequently defeasible as in (2).

What role does this distinction play in children's language comprehension? Recent research has begun to examine how children derive pragmatic inferences that go beyond an initial semantic meaning. For example, when the Child hears Mother's response in (1), can she correctly infer that she has been prohibited from eating the entire cake? Surprisingly, researchers

have found that even school-aged children can be quite literal in their interpretation of utterances, often failing to generate robust inferences (Papafragou & Musolino, 2003; Noveck, 2001; Harris & Pexman, 2003; Vosniadou, 1987; Bernicot, Laval, Chaminaud, 2007; deVilliers, deVilliers, Cole-White, & Carpenter, 2008). One might surmise that children are simply pragmatically incompetent or uninterested in speakers' intentions. But this broad interpretation is difficult to reconcile with the ample evidence that even young toddlers can make sophisticated inferences about the communicative intentions of others during word learning (Tomasello, 1992; Baldwin, 1993).

How then do we make sense of this tension between the pragmatic sophistication of early toddlers and the stubborn adherence to literal meaning by school-aged children? In this paper, we seek to understand the nature of children's surprising failures by exploring their behavior using a more naturalistic measure of comprehension. These experiments recruit children's eye-movements as an implicit indication of their ability to generate post-semantic inferences. Since eye-movements are tightly linked to the processing of spoken language, they can also potentially provide a fine-grained measure of how interpretations unfold over time. Following prior research, we focus on the relationship between word meaning and ultimate interpretation by examining a test case where the division between semantic meaning and pragmatic inference is sharply defined: the interpretation of scalar quantifiers. In the remainder of the Introduction, we will flesh out an account of the semantics of scalars and briefly review recent studies on children's understanding of scalar terms. We will also discuss possible limitations of previous tasks and describe a series of experiments designed to probe the development of pragmatic interpretation through an implicit measure of language processing.

1.1. The Gricean theory of scalar interpretations

Linguists have noted that scalar quantifiers like *some* have two distinct interpretations. Typically, sentences like (3) will imply that the Child ate some but not all of the spinach.

(3)   Mother: Did you eat your spinach?

        Child: I ate some of it.

However, on occasion *some* can be used in a context that does not exclude the total set. For example, Popeye asserts in (4) that he has eaten some of the spinach but then goes on to explain that he ate all of it.

(4)   Olive Oyl: If you ate some of the spinach, I won't have enough for dinner!

        Popeye: I ate some. In fact, I needed me strength, so I ate it all.

Formal treatments of natural language have suggested these two interpretations of *some* are actually the result of a single meaning (Horn, 1972/1989; Gadzar, 1979). As Figure 1 illustrates, the *some* and *all* can be ordered on a scale with respect to the strength of the information that they convey. On this theory, the meaning of the weaker term (*some)* is consistent with all values greater than a lower boundary (*some* is greater than *none*) up through and including the maximum (*all)*. In sentences like (4), this meaning is transparent. Interpretations like this are termed *lower-bounded* since the scalar term has a lower boundary but no upper bound.

INSERT FIGURE 1 HERE

However, *some* is typically interpreted as having an additional boundary which excludes referents which are compatible with *all*. This happens via a pragmatic inference called a *scalar implicature*. According to Grice, the participants in a conversation expect that each will tailor their contribution to be as informative as required but no more informative than is required (Quantity Maxim). To see how this might occur, one can imagine a situation where the Child had actually polished off the spinach and uttered (5).

(5)   Child: I ate all of the spinach.

The existence of this more informative alternative means that if the speaker chooses instead to use a weaker scalar term like in (3), the listener can apply the Quantity Maxim and infer that this was a situation where the stronger scalar term was not true. This interpretation is called *upper-bounded* since it imposes an additional boundary on the upper end of the scale. Thus we can infer that if the Child had eaten all of the spinach, she would have simply said so. But since she did not, she must have eaten only some but not all of it.

This logic can be extended to any set of terms which can be placed on ordinal scale and which differ in their strength (Horn, 1972/1989; Levinson, 2000). Parallel inferences have been noted for a wide range of expressions including scalar adjectives (*warm* vs. *hot*), aspectual verbs (*start* vs. *finish*), and logical operators (*or* vs. *and*). Scalar implicatures can even be generated in cases where alternatives are ordered solely by virtue of the context or our knowledge of common practices (Hirschberg, 1985; Papafragou & Tantalou, 2004; Katsos & Bishop, 2008). For example, in (1) our knowledge of the part-whole relationship establishes a scale where Mother's use of the weaker alternative (slice of cake) leads us to infer that the stronger is not true (entire cake).

While this division between semantic content and pragmatic inference has been widely accepted, there are divergent theories about the nature of these two levels of representation and their relation to one another (Sperber & Wilson, 1986/1995; Recanati, 2003; Levinson, 1983, 2000). More recent Neo-Gricean accounts have argued that the habitual generation of implicatures could result in their automatization and lead to rapid deployment of the restricted meaning during conversation (Levinson, 1983, 2000). Another proposal links the generation of scalar implicatures to the grammatical properties of the sentence (Chierchia, 2004). Finally,

according to the Relevance theory, the calculation of scalar implicatures is associated with the tradeoff between the possible gains associated with generating an inference and the amount of cognitive effort necessary to derive it (Carston, 1998; Recanati, 2003; Sperber & Wilson, 1986/1995). Critically, all these accounts acknowledge that while many aspects of utterances are tightly linked to word meaning and syntactic structure, other facets are clearly added by context-sensitive, inferential processes. This paper explores this dichotomy by testing children's ability to make this simple pragmatic inference during real-time language comprehension.

1.2. Previous research on the development of scalar implicatures

S*ome* and *all* are among the first quantifiers that children produce (Dale & Fenson, 1996). Nonetheless, developmental research has demonstrated profound differences in adult's and children's comprehension of these terms. When presented with a weaker term like *some*, adults consistently favor the upper-bounded reading while children prefer the lower-bounded one (Smith, 1980; Noveck, 2001; Scrafton & Feeney, 2006; Papafragou & Musolino, 2003; Pouscoulous, Noveck, Politzer, & Bastide, 2007; Barner, Chow, & Yang, 2009; Huang, Snedeker, & Spelke, submitted). The same curious pattern emerges across a variety of scalar terms including conjunctions, modals, and predicates (Paris, 1973; Braine & Rumain, 1981; Noveck, 2001; Chierchia, Crain, Guasti, Gualmini, & Meroni, 2001; Papafragou & Musolino, 2003).

While this work seems to suggest that children have a global inability to calculate scalar implicatures, some researchers have argued that these studies vastly underestimate early pragmatic competence (Papafragou & Tantalou, 2004; Papafragou, 2006). In particular, two features of prior tasks may limit the conclusions we can draw. First, most of these studies employ judgment tasks which require children to explicitly reason about another character's

statement. For example, in Papafragou and Musolino (Experiment 1, 2003), adults and children saw a scene where a girl finished a puzzle and heard a puppet utter the statement "The girl started the puzzle." Participants were then asked to evaluate whether this puppet "answered well" in its description of the situation. Thus rather than directly assessing whether the use of *started* led children to infer *not finished*, these tasks instead measured the participants' ability to reason about the felicity of the puppet's use of the weaker term. Such judgments may require significant meta-linguistic awareness, an ability that develops slowly over the school years (Papafragou & Tantalou, 2004; Papafragou, 2006; Pouscoulous et al., 2007).

Second, these judgment paradigms frequently rely on the use of underinformative sentences. For example, in Smith's (1980) original study, adults and preschoolers were asked to agree or disagree with statements like "Some elephants have trunks." Here the weaker term is used to describe situations where all members of the category share a particular property. Similarly, Noveck (2001) asked adults and seven-year-olds whether the puppet's statement "There *might* be a parrot in the box" was correct in a context where they knew that it *must* be so. In both cases, participants' spontaneous judgments revealed their interpretation of the sentence: Rejections indicated an upper-bounded reading while acceptances indicated a lower-bounded one. Notice however that both true and false judgments correspond to semantically valid uses of the quantifier. Thus in order to succeed, participants needed to recognize that the goal of the task relates to assessing the pragmatic felicity of an utterance and not its truth-value. While adults may grasp this subtlety with minimum instruction, children may be less able to do so.

More recent studies have suggested that children can generate scalar implicatures under some circumstances. For example, Papafragou and Musolino (Experiment 2, 2003) encouraged children to assess the felicity of an utterance by telling the child that the speaker says "silly

things" and that their goal is help her "say them better." The children were given several practice trials in which the puppet either labeled things as the child probably would ("a dog") or in an unnatural way ("an animal with four legs"). After this training, children were more likely to reject underinformative statements (*started* when *finished* is true), suggesting that they may have calculated an implicature (48% post-training compared to 10% without training). Similarly, Papafragou (2006), following Papafragou and Tantalou (2004), asked children to reward the puppet for completing particular tasks (building the school). When the puppet subsequently stated "I started to build it," they found that children often correctly withheld the prize (47% of the time). These findings suggest that children are more likely to generate scalar implicatures in situations where the stronger alternative is made more salient.

Recently, Pouscoulous and colleagues (2007) examined whether children are more likely to calculate implicatures when the cognitive demands of doing so are minimized. They used an act out task ("make some boxes contain a token") and varied processing costs by manipulating linguistic and task-related variables. They found that children had less difficulty generating implicatures when less complex scalar expressions were used and when the task involved a smaller number of distractors. Critically, they also observed that children were more likely to interpret *some* with an implicature in this act out task than they had been in prior judgment tasks. Altogether these findings suggest that the degree to which children will generate scalar implicatures varies depending on cognitive load. It also suggests that the developmental difference between children and adults in previous studies on implicature could be driven by children's more limited processing abilities.

This raises the intriguing possibility that if all extraneous task demands were removed— and no action or judgment was required— children's comprehension might be quite similar to

that of adults. In our current study, we examine children's ability to calculate scalar implicatures in a naturalistic task that uses eye-movements as a spontaneous measure to track interpretation during real-time language processing. Unlike prior studies, this paradigm removes all overt task demands and uses sentences that are globally unambiguous. Under these circumstances, will the pragmatic skills of children more closely approximate those of adults?

1.3. Constructing a naturalistic method for assessing the interpretation of scalar terms

The following experiments examine children's sentence comprehension using the visual-world eye-tracking paradigm (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). This procedure presents participants with spoken instructions asking them to manipulate objects within a visual reference world while their eye-movements to those objects are measured. This indirect measure of comprehension has several advantages for exploring the development of the semantic and pragmatic interpretations. First, since eye-movements are typically made without conscious reflection, they provide a more implicit measure of comprehension prior to any overt strategic judgments. Second, because eye-movements are rapid, frequent and tightly linked to the processing of spoken language, they can potentially provide a fine-grained measure of how interpretations unfold over time.

In the following experiments, five- and six-year-olds heard stories in which two types of objects were divided up between four characters. These stories were accompanied by a visual display. In Experiment 1, the items were always divided such that one of the critical characters (the girls) had a proper subset of one item (the socks) while the other had the total set of second item (the soccer balls). In the critical condition, children were given instructions like "Point to the girl that has some of the socks" and their eye-movements were recorded. These trials contained a period of semantic ambiguity beginning at the onset of the quantifier during which

the referent of a lower-bounded reading of *some* is compatible with both characters. Nevertheless, at any point during this time, a scalar implicature can be generated to rule out the girl with the total set (the Distractor) in favor of the girl with the subset (the Target).

We compared eye-movements to the Target in this *some* condition to those in trials asking for "all of the socks" (when the Target has all the socks and the Distractor has a subset of the soccer balls). Since the Distractor in the *all* trials is inconsistent with the semantics of the quantifier, we would predict quick referential disambiguation in these trials. To ensure that differences between these trials were not simply due to preferences for larger quantities or greater difficulties in calculating upper bounds relative to lower bounds, we also used terms from a number scale. Like *all*, *two* and *three* do not require a pragmatic inference to specify exact quantities and thus do not have the same temporary semantic ambiguity as *some* (Papafragou & Musolino, 2003; Huang et al., submitted). Performance on the *two* trials provides a particularly crucial comparison since its meaning rules out the same Distractor as *some* would once the implicature is calculated. By comparing these trials we can examine whether children can restrict reference via semantic content as well as pragmatic implicature.

2. Experiment 1

2.1. Methods

2.1.1. Participants

Twenty adults and 24 five-year-olds (ranging from 5;2 to 6;1, mean age 5;7) participated in this study. For all three experiments, the adult participants were undergraduates at Harvard University who received course credit for their participation. Their performance served as a comparison for children's abilities in this task and full presentation of their data in Experiments 1 and 2 appears in Huang and Snedeker (2009). The children were recruited from the database of

the Laboratory for Developmental Studies which enrolled participants through birth records and the distribution of pamphlets in the Boston area. While information on participants' ethnicity, parental education, income, and occupation was never recorded, prior work with this population suggest that participants predominantly come from middle and high SES homes and are primarily Caucasian. This age group was selected because much of the previous work on children's calculation of scalar implicatures has targeted this age range. All participants in every experiment were native monolingual English speakers.

2.1.2. Procedure

Participants sat in front of an inclined podium divided into four quadrants each containing a shelf where pictures could be placed. A camera at the center of the display was focused on their face and recorded the direction of their gaze while they were performing the task. A second camera recorded both their actions and the location of the items in the display. At the beginning of the study, the experimenter took out pictures of four characters, placed them on each shelf in a pre-specified order, and told the participant that "these boys and girls would receive different things during the game." Every trial consisted of the experimenter acting out a scripted story using pictures of the relevant objects. This was followed by an utterance which instructed participants to pick up a particular character. Once the participant did this, the trial ended, the objects were removed from the display and the next trial began.

2.1.3. Materials

The four quantifiers represented the cells of 2 x 2 design in which the first factor, Quantifier Scale, contrasts terms derived from the critical Gricean scale (*some* and *all*) with terms from the control number scale (*two* and *three*). The second factor, Quantifier Strength, contrasts the weaker quantifiers (*some* and *two*) from the stronger ones (*all* and *three*).

The visual displays featured four characters that were aligned in the following clockwise order beginning from the upper-left quadrant: Craig, Judy, Cheryl, and Pat. This arrangement ensured that the vertically adjacent characters matched in gender while the horizontally adjacent characters did not (see Figure 2). We constructed 16 stories like (6) below. In each story, two types of objects were introduced and distributed among the pairs of boys and girls.

(6) The boys and girls on the soccer team were getting socks and soccer balls from the coach. The coach gave socks to Judy and socks to Craig (*experimenter places 2 socks next to the girl on the upper-right and 2 socks next to the boy on the upper-left*). The coach knew that Pat was already a very good soccer player but he thought that Cheryl needed a lot of practice (*experimenter places a blank card next to the boy on the lower-left and 3 soccer balls next to the girl on the lower-right*).

The distribution of objects among the characters differed according to the Quantifier Scale. For *some* and *all* trials, one set of four items was split evenly between a horizontally adjacent pair (girl with 2 socks and boy with 2 socks) and another set of three items which was given to one child from the remaining pair (girl with 3 soccer balls and boy with no soccer balls). For *two* and *three* trials, the first set was again evenly split between one boy-girl pair while the second set now included a fourth item given to the character who had previously received nothing (boy with 1 soccer ball). This difference between the number and quantifier trials was necessary to ensure that the verbal descriptions were felicitous for all trial types. While saying "three of the socks" would be odd when there are only three socks in total, adding an extra object to the character of opposite gender makes the utterance felicitous without changing the visual properties of the critical Target or Distractor characters (Huang & Snedeker, 2009).

Introducing the objects as part of a single large set and then dividing that among the

characters established a frame of reference that constrained the interpretation of the quantified

phrases. For example, "all of the soccer balls" most naturally referred to all the soccer balls that

the coach has rather than all of the soccer balls in the known universe or all of the soccer balls

that Cheryl has. In addition, the stories ensured that children know the labels that we will be

using for each object. These objects were referred to with definite noun phrases (e.g., *the socks*)

or bare plurals (e.g., *socks*) to ensure that children were not primed to associate a particular

subset with the numbers and quantifiers used in the instructions. In a separate task, these

contexts were verified to successfully establish expectations that (1) quantifiers would refer

specifically to the sets in the display, (2) objects would be identified by basic-level labels, and

(3) *some* would be interpreted with a scalar implicature (Huang & Snedeker, 2009).

INSERT FIGURE 2 HERE

For each story we created a quartet of target sentences, like those shown in (7).

(7)   Point to the girl that has <u>some/all/two/three</u> of the socks.

The target sentences in each condition were identical except for the gender of the child that was

requested and the identity of the final word. The gender of the child was linked to the content of

the story: If the set of three objects had been given to a girl, then a girl was requested. The

names of the two items that were distributed always had the same onset (e.g., socks and soccer

balls), thus creating a brief period of ambiguity during which the identity of this noun was

uncertain. A complete list of the materials for all experiments may be obtained from the first

author. Target sentences were recorded by a female actor and the digital waveforms were

examined to ensure that they had a consistent unmarked prosody. The sound files were edited to

equate the lengths of two critical regions: 1) from sentence onset to the gender cue ("Point to

the") and 2) from the onset of the gender cue to the onset of the quantifier ("girl that has"). Four versions of each base item were used to create four presentation lists such that each list contained four items in each condition and that each base item appeared just once in every list.

2.1.4. Coding

Trained research assistants watched videotapes of the participants' actions and coded them based upon selection of one of the four characters. Across all experiments, we only included trials where participants correctly selected the Target in subsequent analyses of eye movements. Approximately 9.6% of child trials were excluded on this basis. Additionally, 0.9% of adult trials and 1.0% of child trials were excluded from further analyses because of experimenter error.

Eye movements were coded by a research assistant who was blind to the location of each object using frame-by-frame viewing of the participants' face on a Sony digital VCR. Each recorded trial began at the onset of the instruction and ended with completion of the corresponding action. Each change in direction of gaze was coded as towards one of the quadrants, at the center, or missing due to looks away from the display or blinking. These missing frames accounted for 2.0% of coded frames in adults and 12.4% in children. Afterwards these looks were then recoded based on their relation to the final instruction: (1) Target; (2) Distractor; (3) Other characters that did not match gender cues. Twenty-five percent of the trials were checked by second coder who confirmed the direction of fixation for 93.6% of coded frames in adults and 97.3% in children. Any disagreements between the two coders were resolved by a third coder.

2.2 Results

Across all experiments, we first conducted a coarse-grained analysis of adult's and children's fixations as the target utterance unfolded. The dependent measure was total looking

time to the Target as a proportion of looking time to the Target and the Distractor. This score ranged from zero (exclusive looks to the Distractor) to one (exclusive looks to the Target). Fixations to the other characters after onset of the gender cue accounted for less than 10% of total looks across all experiments and were not included in subsequent analyses. Each period was analyzed with subjects and items ANOVAs with Quantifier Scale (number vs. scalar) and Quantifier Strength (lesser vs. greater) as within subject and item variables and list/item group as a between subjects and items variable.

Table 1 lists the duration of the five time windows that were analyzed. Each period is shifted 200ms after the relevant marker in the speech stream to account for the time it would take to program a saccadic eye-movement (Allopenna, Magnuson, & Tanenhaus, 1998; Matin, Shao, & Boff, 1993). The first two regions, the Baseline phase ("Point to the") and the Gender phrase ("girl that has"), provide comparisons of looks to the Target and Distractor before the introduction of any quantifier information. Among adults, there were no reliable effects of Quantifier Scale or Strength during these periods (all F's<4.00, all p's>.05). However, among children the Target looks differed across the trial types even prior to the onset of the quantifier (see Figure 3). There was a greater preference to look at the Target in the *three* and *all* trials relative to the *two* and *some* trials, leading to a significant effect of Quantifier Strength in both the Baseline ($F1(1,19)=25.19, p<.01, \eta^2=.57$; $F2(1,15)=13.85, p<.01, \eta^2=.48$) and Gender phases ($F1(1,19)=10.66, p<.01, \eta^2=.35$; $F2(1,15)=8.63, p<.01, \eta^2=.37$). This indicates that children had a bias to look at the character with the larger quantity of items: the person with 3 in the number trials and the person with the total set on the quantifier trials.

<div align="center">INSERT FIGURE 3 HERE</div>

2.2.1. Adults: Post-quantifier saccade analysis

In order to isolate differences that emerged following the onset of the quantifier, we conducted an additional analysis of saccades initiated after quantifier onset. In these saccade analyses, we separated the trials based on the object that the participant was fixating in the previous frame (Target or Distractor) and calculated the probability of switching to the other object following the onset of the quantifier. Analyses of this kind have been used extensively in research on the development of word recognition (Fernald, Pinto, Swingley, Weinberg, & McRoberts, 1998; Swingley & Fernald, 2002; see also Altmann, & Kamide, 2004) and allow us to factor out early differences in fixation patterns by specifically comparing trials on which participants were looking at the same objects when the quantifier began.

We began by examining Target and Distract saccades during the Quantifier phase. This critical period begins from the onset of the quantifier and ends prior to the onset of the disambiguating phoneme ("some/all/two/three of the soc-"). This region averaged 667ms in length. Our initial coarse grained analysis examined all switches that occurred during this time period. Adults exhibited equal looks to the Target (M=32%, SE=4%) and Distractor (M=30%, SE=4%) across all conditions immediately prior to the region, $t1(19)=0.38, p>.70, \eta^2=.01$; $t2(15)=0.29, p>.70, \eta^2=.01$. However following the onset of the quantifier, we found that adults' switches varied across conditions. While switches to the Target exceeded switches to the Distractor in the *two* (M=36%, SE=5% vs. M=17%, SE=4%), *three* (M=37%, SE=6% vs. M=12%, SE=3%), and *all* trials (M=34%, SE=5% vs. M=17%, SE=4%; all p's<.05), they did not differ from each other in the *some* trials (M=17%, SE=4% vs. M=8%, SE=3%; p>.30). This led to a significant Quantifier Scale by Strength interaction during this period, $F1(1,19)=6.34, p<.05, \eta^2=.25$; $F2(1,15)=0.32, p<.60, \eta^2=.02$.

To explore these differences more closely, we calculated the probability of switching to the other object during each 200ms time window following the quantifier. Thus for each time window we redivided the trials on the basis of the object fixated in the prior frame. This allows us to isolate the moment at which new eye-movements to the Target outnumber new eye-movements to the Distractor. During the 400ms time window, we found a robust Target preference in shifts for the *two* (M=17%, SE=4% vs. M=8%, SE=3%), *three* (M=23%, SE=6% vs. M=3%, SE=2%), and *all* trials (M=24%, SE=6% vs. M=6%, SE=3%), demonstrating rapid rejection of the Distractor when it was inconsistent with the semantics of these terms (all p's <.05). In contrast, in the case of *some,* where the semantics of the quantifier was consistent with both characters, the preference for shifting to the Target did not emerge until the 600ms time window (M=25%, SE=6% vs. M=7%, SE=4; $t1(19)=2.38, p<.05, \eta^2=.23$; $t2(15)=0.71, p<.50,$ $\eta^2=.03$). This suggests that the scalar implicature was not available to rule out the Distractor during the earliest stages of processing. Nevertheless, this 600ms time window occurred before the point of phonological disambiguation, demonstrating that adults calculated the scalar implicature online in this task.

2.2.2. Children: Post-quantifier saccade analysis

Turning to the children, we found relatively equal looks to the Target (M=45%, SE=3%) and Distractor (M=38%, SE=2%) across all conditions immediately before the onset of the quantifier, $t1(23)=1.46, p>.15, \eta^2=.08$; $t2(15)=0.70, p>.40, \eta^2=.03$. However Figure 4 illustrates that during the Quantifier phase, Target switches exceeded Distracter switches in the *two* (M=16%, SE=3% vs. M=8%, SE=3%), *three* (M=28%, SE=5% vs. M=12%, SE=3%), and *all* trials (M=28%, SE=4% vs. M=13%, SE=3%; all p's<.05) while in the *some* trials, Distractor switches exceeded Target switches (M=8%, SE=3% vs. M=24%, SE=6%; p<.05). This led to a

significant Quantifier Scale by Strength interaction, $F1(1,23)=5.11, p<.05, \eta^2=.18$; $F2(1,15)=9.63$, $p<.01, \eta^2=.39$. Fine-grained analyses of 200ms time windows revealed a robust Target preference in shifts for the *three* (M=10%, SE=3% vs. M=4%, SE=2%) and *all* trials (M=23%, SE=6% vs. M=4%, SE=2%) in the 400ms time window while *two* trials (M=10%, SE=4% vs. M=1%, SE=.5%) followed in the 600ms time window (all p's<.05). Critically, unlike adults, children's Target preference in the *some* trials did not exceed their Distractor preference until the 1200ms time window (M=32%, SE=6% vs. M=4%, SE=2%; $t1(23)=4.24, p<.05, \eta^2=.44$; $t2(15)=2.80, p<.05, \eta^2=.34$). Since this period occurs well past the point of phonological disambiguation, these results suggest that the scalar implicature was not available during children's online processing.

INSERT FIGURE 4 HERE

2.2.3. Comparisons between adults and children

Finally, we directly compared adult and child saccades following the onset of *some* through a series of ANOVAs with Initial look (Target vs. Distractor) as a within subject variable and Age (adult vs. child) as a between subject variable. During the Quantifier phase, we found adults were more likely to switch their looks to the Target than Distractor while children were more likely to switch their looks to the Distractor than Target. This led to a significant interaction between Age by Initial look, $F1(1,42)=5.37, p<.05, \eta^2=.11$; $F2(1,30)=4.81, p<.05, \eta^2=.13$. A closer examination revealed that this interaction emerged prior to phonological disambiguation, in the 600ms time window ($F1(1,42)=10.02, p<.01, \eta^2=.19$; $F2(1,30)=3.68, p<.10, \eta^2=.11$) and continued following the disambiguating phoneme into the 1200ms time window ($F1(1,42)=6.73, p<.05$, $\eta^2=.14$; $F2(1,30)=16.34, p<.01, \eta^2=.35$). These results demonstrate that while adults are able to

generate a scalar implicature during online processing, children delay their reference restriction until after the disambiguating phoneme.

2.3. Discussion

In Experiment 1, we found that children's reference resolution was strongly affected by the term they heard. Following the onset of the quantifier, we found increased looks to the Target for the *two*, *three*, and *all* trials but not for the *some* trials. These results demonstrate that when lexical semantics is sufficient to identity the Target, disambiguation is quite rapid. However, when semantic analysis is not sufficient, reference resolution is substantially delayed. Critically, children in the *some* trials, unlike adults, fail to show a reliable Target preference until the end of the instruction. This suggests that rather than calculating the implicature, they simply waited until the correct referent was specified by the disambiguating phoneme (i.e. used *-ks* to select *socks* rather than *soccer balls*).

Could the delays that were observed in the *some* conditions be accounted for some process other than scalar implicature? One might argue that the differences between the gaze time patterns for *some* and *two* are attributable to differences in the verification conditions for numbers and scalar quantifiers. The applicability of a number can be verified solely by looking at the set of objects owned by the character in question (*Does the girl have exactly two socks?*). In contrast, to determine whether the upper-bounded reading of *some* applies, a child would have to examine both the set of objects belonging to the character in question and the set of those objects belonging to the adjacent character (*Does the girl have some socks and does the adjacent boy have at least one as well?*). This might require additional processing and perhaps additional eye-movements as well, explaining why looks to the Target were slower for *some* than for *two*.

We think this account is unlikely for a couple of reasons. First, it incorrectly predicts that we should see parallel delays for *all*. Like *some*, the verification conditions for *all* involve the set of objects that the Target has and the set owned by the adjacent character (*Does the girl have some socks and the does the boy have none?*). Yet we found that the referent for *all* was disambiguated as rapidly as the referent of *three*, suggesting that this difference created no measurable delay. Second, the children's eye-movements do not suggest that were overtly verifying the sentences in this way. Looks to the adjacent character were rare in all conditions, accounting for only 9% of total looks after the onset of the quantifier. Critically, the proportion of such fixations did not differ across conditions, suggesting no privileged strategy for the *some* relative to other quantifiers (all p's>.20).

There is however yet another interpretation of these data. Perhaps the delay that we observed is attributable to general difficulties processing in *some* rather than to sluggish use of scalar implicature. Unlike the other terms, *some* lends itself to two readings. The simultaneous activation and competition of both meanings may have resulted in a stalemate that prevented children from interpreting the relative clause and using it to restrict reference. Such a strategy might also lead participants to delay looking to the Target until the arrival of disambiguating phonological information. Experiment 2 tests this hypothesis. We reasoned that if hearing *some* automatically activates multiple competing meanings, then there should be delays even when the Distractor is inconsistent with both the lower- and upper-bounded interpretations (e.g., a girl with no socks or soccer balls). In contrast, if prior delays reflect a failure to generate the implicature, they should disappear when the semantics of the term is sufficient for reference resolution.

As in Experiment 1, children in one set of *some* trials were presented with a girl that had some-but-not-all of the socks and another that had all of the soccer balls. These will be called

"2-referent trials" because there are two referents that are consistent with the semantics of the quantifier. These trials were compared to a second set of *some* trials where children were presented with a girl that had some-but-not-all of the socks and another that had nothing. These will be called "1-referent trials" because there is only one referent that is consistent with the semantics of *some*. Critically, in these trials, the Target can be resolved solely by the semantics of the term rather than by a scalar implicature. Thus if pragmatic processing is delayed relative to semantic processing, then children should be considerably faster at disambiguating the Target in these trials. If, however, competition between the two readings accounts for the slower resolution of the referent of *some*, then this processing delay should still be present in the 1-referent as well as in the 2-referent trials.

3. Experiment 2

3.1. Methods

3.1.1. Participants

Twenty adults and 24 five-year-olds (ranging from 5;5 to 6;9, mean age 6;0) participated in this study. The children were recruited from Roberts Elementary School in a suburb of Boston. While information on participants' ethnicity, parental education, income, and occupation was never recorded, information from the 2000 census in this community suggests that participants predominantly come from middle SES homes and are primarily Caucasian.

3.1.2. Procedure and Materials

The materials compared the interpretation of *some* in two different referential contexts (see Figure 5). In the 1-referent trials, we introduced them to displays that contrasted a subset quantity of one item with its empty set. Participants heard four new stories where a single set of objects were introduced and distributed among the boy-girl pairs like (8) below.

(8)   The boys and girls on the soccer team were getting socks from the coach. The coach

gave socks to Judy and socks to Craig and socks to Pat (*experimenter places 3 socks*

next to the girl on the upper-right, 3 socks next to the boy on the upper-left, and 3

socks next to the boy on the lower-left). But these socks were too big for Cheryl's big

feet (*experimenter places a blank card next to the girl on the lower-right*).

On these trials, three characters evenly shared nine items (girl and two boys with 3 socks) with a

fourth character receiving nothing (girl with no socks). In the 2-referent trials, we again

introduced participants to stories and displays that contrasted a subset quantity of one item with

the total set of another (see Experiment 1). Following each story, participants heard instructions

asking them to "Point to the girl that has some of the socks."

INSERT FIGURE 5 HERE

We also included an equal number of filler trials to prevent children from predicting the

Target prior to the onset of the quantifier. These filler trials used the same displays as the critical

trials above but used quantifiers that were consistent with the Distractor set. For the 2-referent

displays, participants were asked to select the girl that has "all of the socks" and for the 1-

referent displays, they were asked for one with "none of the socks." Like previous experiments,

four items of each type were presented over the course of 16 randomized trials. The presentation

of materials was counterbalanced by creating four lists such that each item appeared just once of

every list and every item appeared in all four conditions across lists.

3.1.3. Coding

Approximately 0.3% of adult trials and 4.2% of child trials were excluded from further

eye-movement analyses because of incorrect action responses. Approximately 0.3% of adult

trials and 0.8% of child trials were also excluded due to experimenter error. Missing frames due

to blinks or looks away accounted for 4.0% of all coded frames in adults and 5.3% in children. There was 94.1% inter-coder reliability in adults and 92.4% in children.

3.2. Results

Initial examination of the proportions of Target looks in adults reveals no significant differences between 1-referent and 2-referent trials during the Baseline and Gender phases (all t's<1.50, all p's>.15). Similarly, Figure 6 illustrates that while Target looks were slightly below chance across both conditions, there were no reliable differences between the two trial types during the Baseline and Gender phases (all t's<1, all p's>.70).

INSERT FIGURE 6 HERE

3.2.1. Adults: Post-quantifier switch analysis

To explore eye-movements generated after the quantifier, we conducted a saccade analysis for the Quantifier phase. Prior to the onset of the quantifier, adults were equally likely to be looking at the Target (M=42%, SE=3%) and Distractor (M=40%, SE=2%) across both conditions, $t_1(19)=0.53, p>.50, \eta^2=.01$; $t_2(15)=0.50, p>.60, \eta^2=.02$. However during the Quantifier phase, the pattern of switches to the Target and the Distractor differed across conditions. Switches to the Target exceeded switches to the Distractor in the 1-referent trials (M=59%, SE=5% vs. M=7%, SE=2%; p<.01), but did not in the 2-referent trials (M=29%, SE=7% vs. M=39%, SE=6%; p>.30). This led to a significant interaction between condition and type of switch, $F_1(1,19)=23.85, p<.01, \eta^2=.56$; $F_2(1,15)=77.40, p<.01, \eta^2=.84$.

Next we explored how this difference emerged over time with a fine-grained analysis. Our goal in this analysis was to understand when the difference between the 1-referent and 2-referent conditions became reliable. In contrast, our goal in Experiment 1 had been to understand when saccades from the Distractor exceeded saccades from the Target. Thus we used a different

analysis for this experiment – a switch analysis. Rather than using a new baseline for each time window, we used a single baseline (the frame before quantifier onset) and for each time window measured the proportion of subjects on Target-initial trials who were now fixating the Distractor and the proportion of subjects on Distractor-initial trials who were now fixating the Target.

In trials where adults were initially looking at the Target, we found fewer switches to the Distractor in the 1-referent trials, presumably because it was inconsistent with the semantics of *some*. In contrast, adults in the 2-referent trials were more likely to switch their looks to the total set which was consistent with the meaning of the quantifier. This led to a significant difference between trial types that began 300ms after quantifier onset (M=13%, SE=7% vs. M=42%, SE=9%; $t1(19)=2.64,p<.05,\eta^2=.27$; $t2(15)=3.42,p<.01,\eta^2=.44$) and continued into the 900ms time window (M=10%, SE=5% vs. M=28%, SE=8%; $t1(19)=2.15,p<.05,\eta^2=.20$; $t2(15)=2.14$, $p<.05,\eta^2=.23$). In trials where adults were initially looking at the Distractor, we found greater switches to the Target item in the 1-referent trials, presumably because the semantics of *some* ruled out the Distractor with the empty set. In contrast, adults in the 2-referent trials were more likely to continue looking at the Distractor with the total set, again demonstrating an initial delay in calculating the scalar implicature. This led to a significant difference between trial types that began 400ms after quantifier onset (M=52%, SE=6% vs. M=30%, SE=9%; $t1(19)=2.37,p<.05$, $\eta^2=.23$; $t2(15)=3.47,p<.01,\eta^2=.44$) and continued into the 1000ms time window (M=87%, SE=5% vs. M=65%, SE=8%; $t1(19)=2.75,p<.05,\eta^2=.28$; $t2(15)=2.30,p<.05,\eta^2=.26$).

3.2.3. Children: Post-quantifier switch analysis

Turning to the children, we found equal looks to the Target (M=38%, SE=2%) and Distractor (M=37%, SE=2%) prior to the onset of the quantifier across both conditions, $t1(23)=0.13,p>.80,\eta^2=.01$; $t2(15)=0.17,p>.80,\eta^2=.01$. Furthermore during the Quantifier phase,

children, like adults, generated more switches to the Target than the Distractor in the 1-referent

trial (M=34%, SE=4% vs. M=18%, SE=4%; p<.05). In contrast, Target and Distractor switches

did not differ from each other in the 2-referent trials (M=18%, SE=4% vs. M=21%, SE=5%; p >

.60), leading to a significant interaction between condition and type of switch in this period,

$F1(1,23)=4.62,p<.05,\eta^2=.17$; $F2(1,15)=1.18,p< .30,\eta^2=.08$.

To understand this interaction, we examined Target and Distractor-initial trials separately

using the switch analysis described above (Figure 7). In the Target-initial trials, there were no

reliable differences between the two conditions. However, on Distractor-initial trials, children in

1-referent contexts began switching to the Target following quantifier onset, while children in 2-

referent contexts were more likely to continue to look at the Distractor. Like the adults, this led

to a significant difference between trial types that began 300ms after quantifier onset (M=43%,

SE=6% vs. M=22%, SE=5%; $t1(23)=2.52,p<.05,\eta^2=.22$; $t2(15)=2.37,p<.05,\eta^2=.27$) and

continued into the 900ms time window (M=73%, SE=5% vs. M=52%, SE=5%; $t1(23)=4.10$,

$p<.01,\eta^2=.42$; $t2(15)=3.08,p<.01,\eta^2=.39$).

INSERT FIGURE 7 HERE

3.2.4. Comparisons between adults and children

Finally, we directly compared changes in fixation in adults and children after the onset of

some by conducting a series of ANOVAs on the switch data with condition (1-referent vs. 2-

referent) as a within subject variable and age (adult vs. child) as a between subject variable. In

the Target-initial trials, adults were less likely to switch to the Distractor in the 1-referent trials

than in the 2-referent trials, while children demonstrated no difference in switches across the two

conditions. This led to a significant difference between conditions ($F1(1,42)=16.15,p<.01$,

$\eta^2=.28$; $F2(1,30)=11.80,p<.01,\eta^2=.28$) as well as an interaction between age and condition

$(F1(1,42)=10.67, p<.01, \eta^2=.20; F2(1,30)=17.41, p<.01, \eta^2=.37)$ which began 200ms after

quantifier onset $(F1(1,42)=6.74, p<.05, \eta^2=.14; F2(1,30)=7.09, p<.05, \eta^2=.19)$ and continued into

the 700ms time window $(F1(1,42)=6.81, p<.05, \eta^2=.14; F2(1,30)=4.37, p<.05, \eta^2=.13)$.

In the Distractor-initial trials, however, we found that both adults and children were more

likely to switch their looks to Target in the 1-referent trials than in the 2-referent trial. This lead

to a significant difference between conditions $(F1(1,42)=15.83, p<.01, \eta^2=.27; F2(1,30)=34.82,$

$p<.01, \eta^2=.54)$ but no interaction between age and condition $(F1(1,42)=1.35, p>.20, \eta^2=.04;$

$F2(1,30)=9.65, p<.01, \eta^2=.24)$. In fact, there were no time windows where adults and children

differed in their switches to the Target across the two conditions (all p's>.15). These results

demonstrate that both adults and children rapidly restrict the referent of *some* when the Distractor

is inconsistent with the semantics of the quantifier.

3.3. Discussion

In Experiment 2, we again found that children were delayed in their looks to the Target for

trials that contrasted *some* with a total set (2-referent trials). However, similar delays were not

seen in trials that contrasted *some* with an empty set (1-referent trials). This pattern was

confirmed when we separated trials by initial fixations and found that following the onset of

some, both adults and children in the 1-referent trials were faster to switch their looks to the

Target but slower in the 2-referent trials. These results suggest that resolution of the Target is

quicker via semantic analysis than pragmatic inference.

However, while these findings suggest that children might altogether fail to generate scalar

implicatures during comprehension, one feature of the data might lead some to be more

skeptical. In both Experiments 1 and 2, children were somewhat slower than adults in using the

control quantifiers to restrict reference, demonstrating a 100 to 200ms delay across the various

conditions. The delay in the case of *some* is clearly greater (600ms in Experiment 1), but we cannot rule out the possibility that children, like adults, calculate the scalar implicature online but do it so slowly that the information only becomes available after phonological disambiguation. In Experiment 3, we explored this possibility by presenting children with situations in which the implicature was explicitly violated by the context. We reasoned that if children were spontaneously generating the inference, then there should be delays when *some* ultimately refers to the total set. In contrast, if they never calculated the inference, then processing in these trials should be no different than trials where *some* refers to the subset.

As in the previous experiments, children in one condition were asked for "some of the socks" in a context where one girl had some-but-not-all of the socks and another had all of the soccer balls. These will be called "SI-consistent trials" because the Target possesses a quantity that is consistent with the scalar implicature. Children in a second condition were asked for "some of the socks" in a context where one girl had all of the socks and another had some-but-not-all of the soccer balls. These will be called "SI-violating trials" because the Target possesses a quantity that violates the scalar implicature. Our critical analyses focused on eye-movements after the disambiguating phoneme. If children are implicitly generating a scalar implicatures, then we would expect greater delays in looks to the Target following the disambiguating phoneme in these trials. However, if children fail to generate an implicature, we would expect that latency to the Target would be the same in these two trials.

4. Experiment 3

4.1. Methods

4.1.1. Participants

Twenty adults and 24 five-year-olds (ranging from 5;6 to 6;8, mean age 6;0) participated in this study. The data from these adults have not been previously published. The children were recruited from Columbus Elementary School in a suburb of Boston and their demographics match those of participants in Experiment 2.

4.1.2. Procedure and Materials

The materials compared the interpretation of *some* in two different referential contexts by adopting displays, stories, and instructions similar to the scalar trials in Experiment 1 (Figure 8). One set of three items was given to one character from the first pair (girl with 3 socks and boy with no socks) and another set of six items was split evenly between the remaining pair (girl with 3 soccer balls and boy with 3 soccer balls). In the SI-violating trials, participants heard instructions asking them to select the Target with the total set, e.g., "Point to the girl that has some of the socks." In contrast, in the SI-consistent trials, the instructions requested the Target with the subset. We also included 16 filler trials to prevent predictability of the Target prior to the onset of the quantifier. These trials used the same displays as the ones described above but used different quantifiers to describe the various sets, e.g., "all/three/none of the socks."

INSERT FIGURE 8 HERE

4.1.3. Coding

Approximately 0.8% of adult trials and 1.7% of child trials were excluded from further eye-movement analyses because of incorrect actions. Approximately 1.2% of adult trials and 1.7% of child trials were excluded due to experimenter error. Finally, missing frames due to blinks or looks away accounted for 3.8% of all coded frames in adults and 4.4% in children. Inter-coder reliability was 94.1% in adults and 93.4% in children.

4.2. Results and Discussion

In adults, the proportions of Target looks did not differ between SI-consistent and SI-violating trials during the Baseline, Gender, and Quantifier phases (all t's<1.50, all p's>.15). Figure 9 illustrates that in the children, however, there was a reliable effect of trial in the Gender phase, Target fixations in the SI-violating trials were greater than in the SI-consistent trials, $t1(23)=2.80, p<.05, \eta^2=.25$; $t2(15)=3.99, p< .01, \eta^2=.51$. This pattern is similar to the bias seen in the initial periods of Experiment 1; in both cases children preferred to look at the character with the unique set (the girl with all the soccer balls). Nevertheless during the Quantifier phase, this difference across the two trial types disappeared ($p>.20$).

INSERT FIGURE 9 HERE

4.2.1. Adults: Post-disambiguation switch analyses

In this experiment the critical time window was the disambiguation phase which begins at the onset of the disambiguating phoneme and ends at the offset of the command ("-ks"). This region averaged 467ms in length. On the frame immediately before disambiguation, adults exhibited equal looks to the Target (M=48%, SE=3%) and Distractor (M=46%, SE=3%) across both conditions, $t1(19)=0.21, p>.80, \eta^2=.01$; $t2(11)=0.13, p>.80, \eta^2=.01$. However, during the Disambiguation phase, we found that the switch patterns varied across conditions. While switches to the Target exceeded switches to the Distractor in both the SI-consistent (M=64%, SE=6% vs. M=2%, SE=1%) and the SI-violating trials (M=49%, SE=4% vs. M=11%, SE=4%; all p's<.05), they did so to a greater degree in the consistent trials. This led to a significant interaction between condition and type of switch during this period, $F1(1,19)=35.40, p< .01$, $\eta^2=.65$; $F2(1,11)=3.94, p<.10, \eta^2=.27$. These results suggest that adults had generated the scalar implicature and were having greater difficulty integrating the disambiguating information when it violated the implicature.

We explored this difference in greater detail using the fine-grained switch analysis from Experiment 2. On Target-initial trials, adults generated more switches to the Distractor in the SI-violating trials, presumably because the Target with the total set violated the pragmatically felicitous use of *some*. In contrast, adults in the SI-consistent trials were more likely to continue looking at the Target with the subset which was consistent with the inferred meaning of *some*. This led to a significant difference between trial types that began 600ms after disambiguation (M=3%, SE=2% vs. M=16%, SE=5%; $t_1(19)=2.11$, $p<.05$, $\eta^2=.19$; $t_2(11)=0.47$, $p<.70$, $\eta^2=.02$) and continued into the 800ms time window (M=1%, SE=.01% vs. M=6%, SE=3%; $t_1(19)=2.17$, $p<.05$, $\eta^2=.20$; $t_2(11)=1.55$, $p<.15$, $\eta^2=.18$). Similarly, on Distractor-initial trials, adults generated more switches to the Target in the SI-consistent trials than in the SI-violating trials. This led to a significant difference between trial types that began 200ms after disambiguation (M=35%, SE=7% vs. M=16%, SE=4%; $t_1(19)=2.70$, $p<.05$, $\eta^2=.28$; $t_2(11)=0.60$, $p<.60$, $\eta^2=.03$) and continued into the 400ms time window (M=75%, SE=6% vs. M=60%, SE=6%; $t_1(19)=2.15$, $p<.05$, $\eta^2=.19$; $t_2(11)=0.52$, $p<.60$, $\eta^2=.02$).

4.2.2. Children: Post-disambiguation switch analyses

Prior to the disambiguating phoneme, children looked equally at the Target (M=44%, SE=2%) and the Distractor (M=43%, SE=2%) across the two conditions, $t_1(23)=0.21$, $p>.80$, $\eta^2=.01$; $t_2(11)=0.13$, $p>.80$, $\eta^2=.01$. During the Disambiguation phase, children, like adults, generated more switches to the Target than Distractor in both the SI-consistent (M=25%, SE=4% vs. M=7%, SE=3%) and the SI-violating trials (M=30%, SE=4% vs. M=7%, SE=3%; all p's<.05). However Figure 10 illustrates that unlike adults, children were equally likely to generate new looks to the Target in both conditions. This led to a significant difference in type of switch ($F_1(1,23)=28.79$, $p<.01$, $\eta^2=.56$; $F_2(1,11)=6.15$, $p<.05$, $\eta^2=.36$) but no critical interaction

between condition and switch type ($F1(1,23)=0.27,p>.60,\eta^2=.01$; $F2(1,11)=1.25,p<.30,\eta^2=.09$). In fact, there were no time windows on either the Target-initial trials or the Distractor-initial trials where switches differed between the two conditions (all p's>.15). Thus we found no evidence that children were slower in their reference resolution in the SI-violating trials even after the onset of the disambiguating phoneme. All together, these results suggest that they failed to calculate the inference during comprehension.

INSERT FIGURE 10 HERE

4.2.3. Comparisons between adults and children

To directly compare the adults and children we entered the switch data for each time window an ANOVA with condition (SI-consistent vs. SI-violating) as a within-subject variable and age (adult vs. child) as a between-subject variable. On the Target-initial trials, adults were more likely to switch their looks to the Distractor in the SI-violating trials than in the SI-consistent trials, while children demonstrated no difference in switches across the two conditions. This lead to a significant interaction between age and condition during the Disambiguation phase ($F1(1,42)=4.11,p<.05,\eta^2=.09$; $F2(1,22)=1.65,p<.30,\eta^2=.07$) that began 500ms after disambiguation ($F1(1,42)=6.66,p<.05,\eta^2=.14$; $F2(1,22)=0.92,p<.40,\eta^2=.04$) and continued into the 800ms time window ($F1(1,42)=4.15,p<.05,\eta^2=.09$; $F2(1,22)=0.31,p<.60,$ $\eta^2=.01$). On the Distractor-initial trials, we found that adults were more likely to switch their looks to the Target in the SI-consistent trials than in the SI-violating trials while children again demonstrated no difference in switches across the two conditions. This lead to another significant interaction between age and condition during the Disambiguation phase ($F1(1,42)=5.92,p<.05,$ $\eta^2=.12$;$F2(1,22)=15.81,p<.01,\eta^2=.42$) which began 200ms after disambiguation ($F1(1,42)=6.56,$ $p<.05,\eta^2=.14$; $F2(1,22)=1.43,p<.30,\eta^2=.06$) and continued into the 400ms time window

$(F1(1,42)=4.19, p<.05, \eta^2=.09; F2(1,22)=2.42, p<.20, \eta^2=.10)$. Altogether these findings demonstrate that while adults were slower to resolve the referent of *some* when it violated the implicature, children show no sensitivity to this pragmatic infelicity.

5. General Discussion

This study explores the development of semantic and pragmatic interpretation by examining children's ability to generate scalar implicatures. In Experiment 1, we found quick resolution of the referent when adults and children hear *two*, *three*, and *all* but delays when they hear *some*. Critically, while adults eventually generate the scalar implicature and use it to restrict reference, children fail to generate it, relying instead on the phonological disambiguation to determine reference. Experiment 2 demonstrated that this delay occurs only when reference restriction requires an implicature and not when semantic analysis is sufficient. Finally, in Experiment 3, we found that children, but not adults, fail to distinguish between contexts that are consistent with the implicature and those that violate it. All together, these results indicate that children do not calculate scalar implicatures during online language comprehension. Instead, their moment-to-moment interpretation of quantified phrases appears to be predominately guided by the semantics of these phrases.

In the remainder of this discussion, we focus on two additional issues. First, we integrate our findings with the existing literature on children's generation of scalar implicatures. Second, we return to the puzzle of why children appear to exhibit pragmatic sophistication in early word learning but fail to generate pragmatic inferences in the case of scalar implicatures.

5.1. Why don't children calculate scalar implicatures?

Our findings demonstrate that children as well as adults initially interpreted *some* with respect to its lower-bounded semantics. Critically, while adults calculated a scalar implicature to

exclude referents compatible with *all* prior to the disambiguating phoneme, children never invoke this late inferential process. Instead, they strictly adhered to the semantics of the quantifier. These findings add to a growing literature demonstrating that children rely heavily on the logical meaning of utterances and have only a limited ability to generate post-semantic inferences (Smith, 1980; Noveck, 2001; Papafragou & Musolino, 2003). Our data offer two new insights into this limitation. First, we demonstrate that this difficulty persists in a naturalistic task requiring no overt judgment or action. Second, by revealing how interpretation evolves over time, our findings demonstrate developmental continuity in the initial semantic processes, followed by developmental discontinuity in subsequent pragmatic processing.

So what accounts for children's failure to make implicatures? The gradual pattern of acquisition and the partial success of young children in some tasks, suggests their failures cannot be attributed to global ignorance of the process (see Introduction). Several other possibilities have been suggested in the literature. One hypothesis, suggested by Relevance Theory, is that children simply have lower thresholds for the relevance of an utterance and thus they do not need to engage in the effortful process of generating this inference (Sperber & Wilson, 1986/1995; Carston, 1998; Recanati, 2003). A second hypothesis suggests that the generation of scalar implicatures is linked to an ability to engage in controlled analytic reasoning that is not possessed by children (Feeney, Scrafton, Duckworth, & Handley, 2004; Scrafton & Feeney, 2006; De Neys & Schaeken, 2007). However, neither account clearly describes the nature of the computational problem that makes scalar implicatures so difficult they require this extra effort or controlled processing.

We see two reasons why these inferences may be particularly tricky for children. First, the contexts of language use typically do not provide evidence that the implicature is necessary.

Any situation compatible with the upper-bounded reading of *some* (e.g., cases in which not all cookies were eaten) is also consistent with the lower-bounded reading (e.g., possibly all cookies were eaten), creating a subset problem. Thus once the child has settled on a lower-bounded meaning, individual examples of this kind cannot demonstrate that the implicature is necessary. A second possibility is that children fail to make implicatures because they are less likely to retrieve the stronger scalar alternative during language comprehension (Gualmini, Crain, Meroni, Chierchia, & Guasti, 2001; Reinhart, 1999; Pouscoulous et al., 2007). On most theories the calculation of the implicature is prompted by awareness of a more informative alternative (for *some*, the term *all*). If this alternative is not retrieved, there is no reason to make the inference. This might explain why children were much more likely to generate the inference in tasks that highlight the need for the stronger term (Papafragou & Tantalou, 2004; Papafragou, 2006).

Nevertheless, it is difficult to see how a failure to retrieve the contrasting term could fully account for children's failure to generate implicatures in the present task. In these experiments, participants saw the subset explicitly contrasted with the total set in both the story and the visual displays as well as in the use of the critical term (*all*) in target utterances throughout the task. This suggests another hypothesis. Remember that we found that adults demonstrated a lag between semantic processing and the calculation of the pragmatic inference. This delay is consistent with adult studies on the processing of scalar implicatures (Bott & Noveck, 2004; Breheny, Katsos, & Williams, 2006; Huang & Snedeker, 2009; Huang & Snedeker, submitted), and suggests a possible linkage between the sluggish processing of these inferences in adults and their sluggish development in children (Gualmini et al., 2001; Reinhart, 1999; Pouscoulous et al., 2007). In particular, it may be the case that the procedures which require more cognitive resources in adults are in turn less likely to be available in children.

5.2. Bottom-up vs. top-down pragmatic processes

In the Introduction we noted a curious disconnect in the literature on the development of pragmatic competence. Toddlers are frequently depicted as pragmatic sophisticates who engage in a rich interpretation of speakers' communicative intentions to learn novel words (Tomasello, 1992/1999; Baldwin, 1991/1993). In contrast school-aged children are surprisingly poor at going beyond the literal meaning of an utterance to infer the speakers intended meaning (Papafragou & Musolino, 2003; Noveck, 2001; Harris & Pexman, 2003; Vosniadou, 1987; Bernicot et al., 2007; deVilliers et al., 2008). What accounts for this apparent developmental discontinuity?

One possibility lies in the inherent asymmetry between the roles of pragmatics in word learning as compared to scalar implicatures. In a typical word learning scenario, the child can directly infer the meaning of a novel word like *doggie* based on a non-linguistic analysis of the speaker's intentions (i.e. seeing the mother point to or look at the family pet). Here pragmatics plays a top-down role in constraining the range of candidates for the meaning of a lexical item. Critically, in this situation, the pragmatic processes can begin independent of and prior to any linguistic semantic analysis. In the case of scalar implicature however, the child can only generate the relevant implicature after some degree of semantic analysis is completed. Specifically, in the present case the child must identify the word *some* and access its meaning before he can evaluate its informational strength relative to *all* and make the upper-bounding inference. Thus the scalar inference depends on the coordination of semantic and pragmatic processes in real time.

This distinction between top-down inferences about communicative intentions and the inferences which build off of bottom-up semantic processes, appears to correctly predict which tasks will be easy for toddlers, and which will be tricky for school-agers. In addition to their

failures with scalar implicature, children struggle with irony, metaphor, and relevance implicatures (Harris & Pexman, 2003; Vosniadou, 1987; Bernicot et al., 2007; deVilliers et al., 2008). In each case development is gradual, variable and prolonged. In each case, pragmatic success requires listeners to calculate an interpretation that builds upon but goes beyond the initial linguistic meaning. These post-semantic processes may be particularly difficult because they require that some feature of the child's initial analysis be revised. This may be another manifestation of the general difficulty that children have in over-riding an initial misinterpretation, parallel to the difficulties that they have in revising syntactic garden paths (Trueswell, Sekerina, Hill, & Logrip, 1999; Snedeker & Trueswell, 2004), accessing the dispreferred meaning of a homonym (Mazzocco, 1997; Doherty, 2004), or identifying a word relative in the presence of competitors (Arnold, 2008; Huang & Snedeker, submitted).

What changes over the course of development to make implicature generation more robust? Preliminary cross-sectional work suggests that children gradually begin to reliably generate scalar implicatures during middle childhood and early adolescence (Guasti et al., 2005; Feeney et al., 2004; Scrafton & Feeney, 2006). This period is characterized by substantial improvements in cognitive control mechanisms which have been implicated in the ability to revise garden-path sentences and other default interpretations (Novick, Trueswell, & Thompson-Schill, 2005; Trueswell et al., 1999). In the case of scalar implicatures, these changes in the language processing system may help children to revise their initial semantic interpretation or may increase the salience of stronger terms on the scale during real-time comprehension. Future studies focusing on comparisons across age groups may permit us to explore these possibilities more deeply.

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38*, 419-439.

Altmann, G., & Kamide, Y. (2004). Now you see it, now you don't: Mediating the mapping between language and the visual world. In: J. Henderson & F. Ferreira (Eds.). *The Interface of Language, Vision, and Action*, 347-386. New York: Psychology Press.

Arnold, J. (2008). THE BACON not the bacon: How children and adults understand accented and unaccented noun phrases. *Cognition*, 108, 69-99.

Baldwin, D. A. (1993). Infants' ability to consult the speaker for clues to word reference. *Journal of Child Language*, 20, 395-418.

Barner, D., Chow, K., & Yang, S. (2009). Finding one's meaning: A test of the relation between quantifiers and integers in language development. *Cognitive Psychology*, 58, 195-219.

Bernicot, J., Laval, V., & Chaminaud, S. (2007). Nonliteral language forms in children: In what order are they acquired in pragmatics and metapragmatics? *Journal of Pragmatics*, 39, 2115-2132.

Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language*. 51(3), 437-457.

Braine, M., & Rumain, B. (1981). Children's comprehension of "or": evidence for a sequence of competencies. *Journal of Experimental Child Psychology*, 31, 46–70.

Breheny, R., Katsos, N., & Williams, J. (2006). Are generalized scalar implicatures generated by default? An on-line investigation into the role of context in generating pragmatic inferences. *Cognition*, 100, 434-463.

Carston, R. (1998). Informativeness, relevance and scalar implicature. In: R. Carston & S. Uchida (eds.). *Relevance Theory: Applications and Implications*, 179-236. Amsterdam: John Benjamins.

Chierchia, G., Crain, S., Guasti, M.T., Gualmini, A., & Meroni, L. (2001). The acquisition of disjunction: Evidence for a grammatical view of scalar implicatures. In A. H.-J. Do, L. Domingues, & A. Johansen, (Eds.), Proceedings of the 25th Boston University Conference on Language Development (pp. 157-168). Somerville, MA: Cascadilla Press.

Chierchia, G. (2004). Scalar implicatures, polarity phenomena, and the syntax/pragmatic interface. In A. Belletti (Ed.)*, Belletti structures and beyond*. Oxford: Oxford University Press.

Dale, P. S. & Fenson, L. (1996). Lexical development norms for young children. *Behavior Research Methods, Instruments, and Computers*, 28, 125-127.

deVilliers, P., deVilliers, J., Cole-White, D., & Carpenter, L. (2008). Acquisition of relevance implicatures in typically-developing children and children with autism. Paper presented at the 33rd annual meeting of the Boston University Conference on Language Development in Boston, MA.

Doherty, M. J. (2004). Children's difficulty in learning homonyms. *Journal of Child Language,*
31, 203-214.

De Neys, W. & Schaeken, W. (2007). When people are more logical under cognitive load: Dual task impact on scalar implicature. *Experimental Psychology, 54*, 128-133.

Feeney, A., Scafton, S., Duckworth, A., & Handley, S. J. (2004). The story of some: Everyday pragmatic inferences by children and adults. *Canadian Journal of Experimental*

Psychology, 58, 121-132.

Fernald, A., Pinto, J., Swingley, D., Weinberg, A., & McRoberts, G. (1998). Rapid gains in

speed of verbal processing by infants in the second year.*Psychological Science*, 9, 228-231.

Gadzar, G. (1979). *Pragmatics: Implicature, presupposition and logical form*. New York:

Academic Press.

Grice, H. P. (1957). Meaning. *Philosophical Review*, 66, 377-88.

Grice, H. P. (1975). Logic and Conversation. In P. Cole and J.L. Morgan (Eds.), *Syntax and

Semantics*, Vol. 3. (41-58). New York: Academic Press.

Gualmini, A., Crain, S., Meroni, L., Chierchia, G., & Guasti, M. T. (2001). At the

semantics/pragmatics interface in child language. *Proceedings of Semantics and

Linguistic Theory XI*. Ithaca, NY: CLC Publications.

Guasti, M. T., Chierchia, G., Crain, S., Foppolo, F., Gualmini., A, & Meroni, L. (2005). Why

Children and Adults Sometimes (but not always) Compute Implicatures. *Language and

Cognitive Processes*, 20, 667-696.

Harris, M., & Pexman, P. M. (2003). Children's perceptions of the social functions of verbal

irony. *Discourse Processes, 36,* 147-165.

Hirschberg J. (1985). *A theory of scalar implicature*. Doctoral diss., University of Pennsylvania.

Horn, L. (1972). *On the semantic properties of the logical operators in English*. Doctoral

dissertation, UCLA, Los Angeles, CA. Distributed by IULC, Indiana University,

Bloomington, IN.

Horn, L. (1989). *A natural history of negation*. Chicago, IL: University of Chicago Press.

Huang, Y., Snedeker, J., & Spelke, E. (submitted). What exactly do numbers mean?

Huang, Y. & Snedeker, J. (2009). On-line Interpretation of Scalar Quantifiers: Insight

into the Semantic-Pragmatics Interface. *Cognitive Psychology*, 58, 376-415.

Huang, Y. & Snedeker, J. (submitted). *Logic and Conversation* revisited: Evidence for a

division between semantic and pragmatic content in real time language comprehension.

Huang, Y. & Snedeker, J. (submitted). Cascading activation across levels of representation in

children's lexical processing.

Katsos, N. & Bishop, D. (2008). A developmental investigation of the effects of scale type and

speech-act on the generation of scalar implicatures. A paper presented at the 11th Congress

of the International Association for the Study of Child Language. Edinburgh, Scotland.

Levinson, S. (1983). *Pragmatics*. Cambridge: Cambridge University Press.

Levinson, S. (2000). *Presumptive meanings*. Cambridge, MA: MIT Press.

Matin, E., Shao, K.C. & Boff, K.R. (1993) Saccadic overhead: information processing time with

and without saccades. *Perception & Psychophysics*, 53(4), 372-380.

Mazzocco, M. M. M. (1997). Children's interpretations of homonyms: A developmental

study.

Journal of Child Language, 24, 441-467.

Noveck, I. A. (2001). When children are more logical than adults: experimental investigation of

scalar implicatures. *Cognition*, 78, 165-188.

Novick, J.M., Trueswell, J.C., & Thompson-Schill, S.L. (2005). Cognitive control and parsing:

Re-examining the role of Broca's area in sentence comprehension. *Journal of Cognitive,*

Affective, and Behavioral Neuroscience, 5(3), 263-281.

Papafragou, A. & Musolino, J. (2003). Scalar implicatures: experiments at the semantics-

pragmatics interface. *Cognition*, 86, 253-282.

Papafragou, A. & Tantalou, N. (2004). Children's computation of implicatures. *Language*

Acquisition, 12, 71-82.

Papafragou, A. (2006). From scalar semantics to implicature: Children's interpretation of aspectuals. *Journal of Child Language*, 33, 721-757.

Pouscoulous, N., Noveck, I. A., Politzer, G., & Bastide, A. (2007). A developmental investigation of processing costs in implicature production. *Language Acquisition*, 14 (4), 347-375.

Paris, S. (1973). Comprehension of language connectives and propositional logical relationships. *Journal of Experimental Child Psychology*, 16, 278–291.

Recanati, F. (2003). *Literal Meaning*. Cambridge: Cambridge University Press.

Reinhart, T. (1999). The processing cost of reference-set computation: Guess patterns in acquisition. OTS Working Papers (Uil-Ots 99001-CL/TL).

Scrafton, S. & Feeney, A. (2006). Dual processes, development, and scalar implicature. Paper presented at the 28th Annual Conference of the Cognitive Science Society.

Smith, C. L. (1980). Quantifiers and question answering in young children. *Journal of Experimental Child Psychology*, 30, 191-205.

Snedeker, J. & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology*, 49, 238-299.

Sperber, D. & Wilson, D. (1986/1995). *Relevance: Communication and cognition*. Oxford: Blackwell.

Swingley, D. & Fernald, A. (2002). Recognition of words referring to present and absent objects by 24-month-olds. *Journal of Memory and Language*, 46, 39-56.

Tanenhaus, M. K., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of

visual and linguistic information in spoken language comprehension. *Science*, 268, 1632.

Tomasello, M. (1992). The social bases of language development. *Social Development*, 1, 67-87.

Trueswell, J.C., Sekerina, I., Hill, N.M. & Logrip, M.L. (1999). The kindergarten path effect:

studying on-line sentence processing in young children. *Cognition*, 73, 89-134.

Vosniadou, S. (1987). Children and metaphors. *Child Development*, 58, 870-885.

Figure 1: Two interpretations of *some*. (A) LOWER-BOUNDED READING: The semantics of *some* can be described as referring to a ray along a quantity scale. (B) UPPER-BOUNDED READING: *Some* is typically interpreted with a pragmatic inference that excludes *all*.

(A)

SOME   ALL

(B)

SOME   ALL
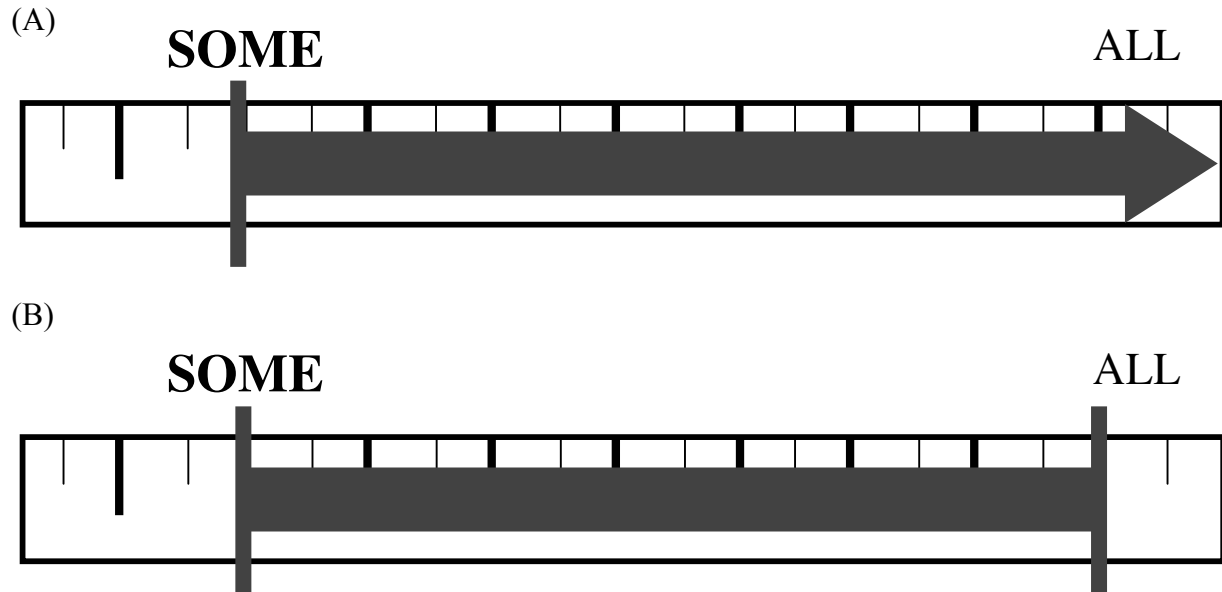
Figure 2: In Experiment 1, examples of visual-world displays for (A) *some*, (B) *two*, (C) *all*, and (D) *three* trials. Participants here were instructed to "Point to the girl that has _____ of the socks." The girl with socks was the Target while the girl with soccer balls was the Distractor.

(A)

(B)

(C)

(D)
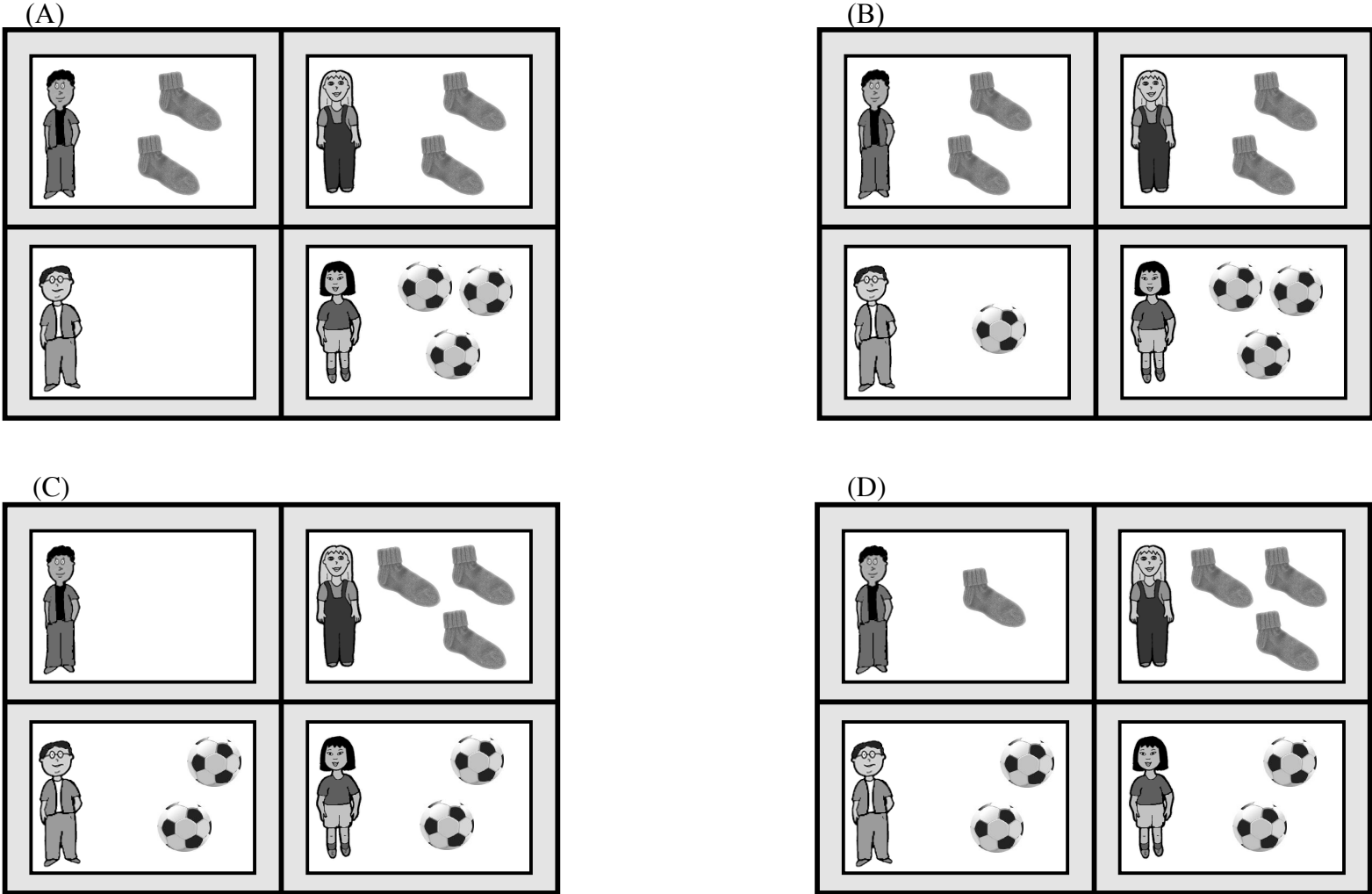
Figure 3: Experiment 1, children. Looks to the Target for the four trial types in coarse-grained time windows.
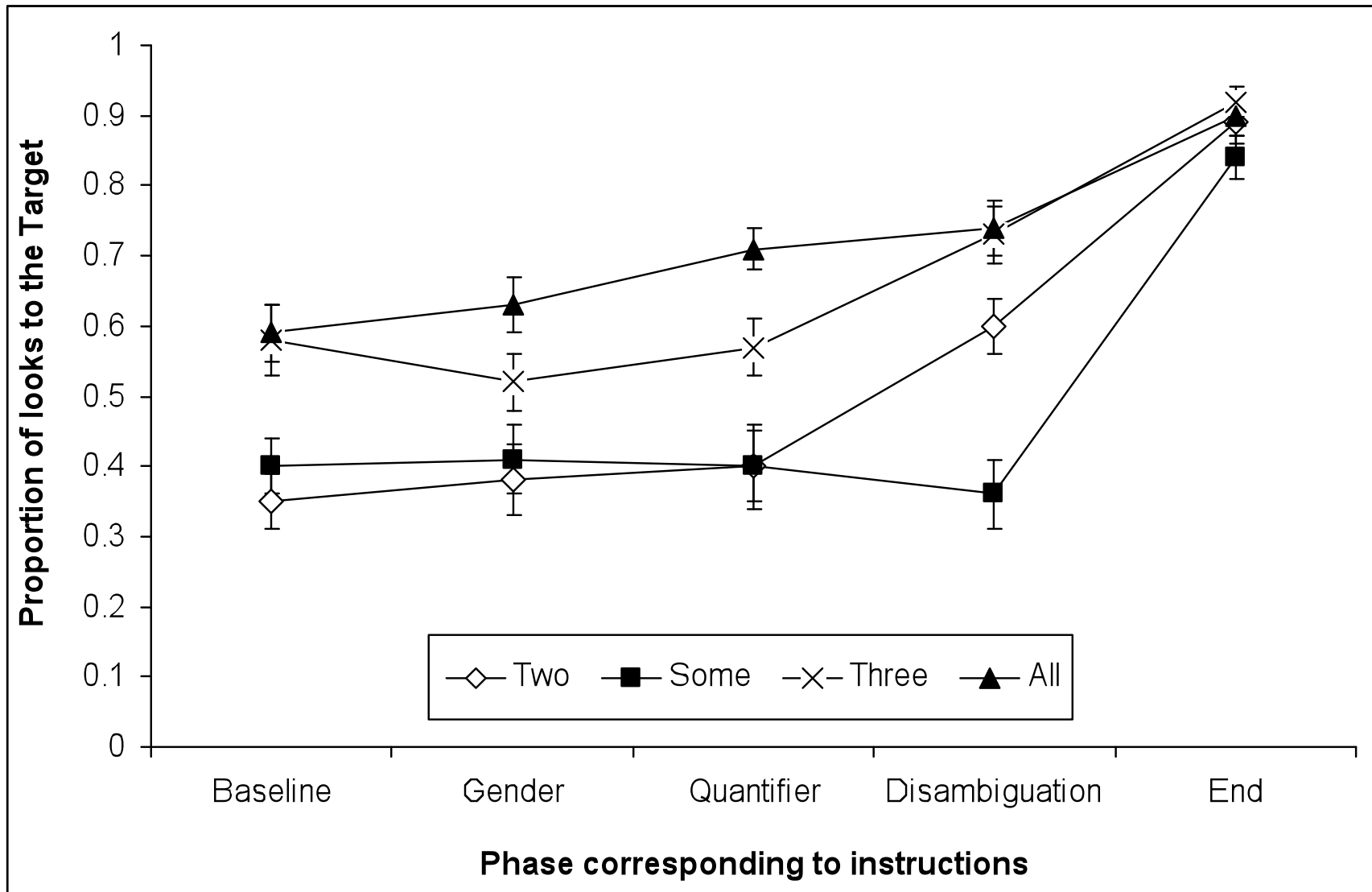
Figure 4: Experiment 1, children. Trials were separated based on fixations prior to the onset of the quantifier (A) Distractor-initial trials: proportion of switches to the Target and (B) Target-initial trials: proportion of switches off the Target.
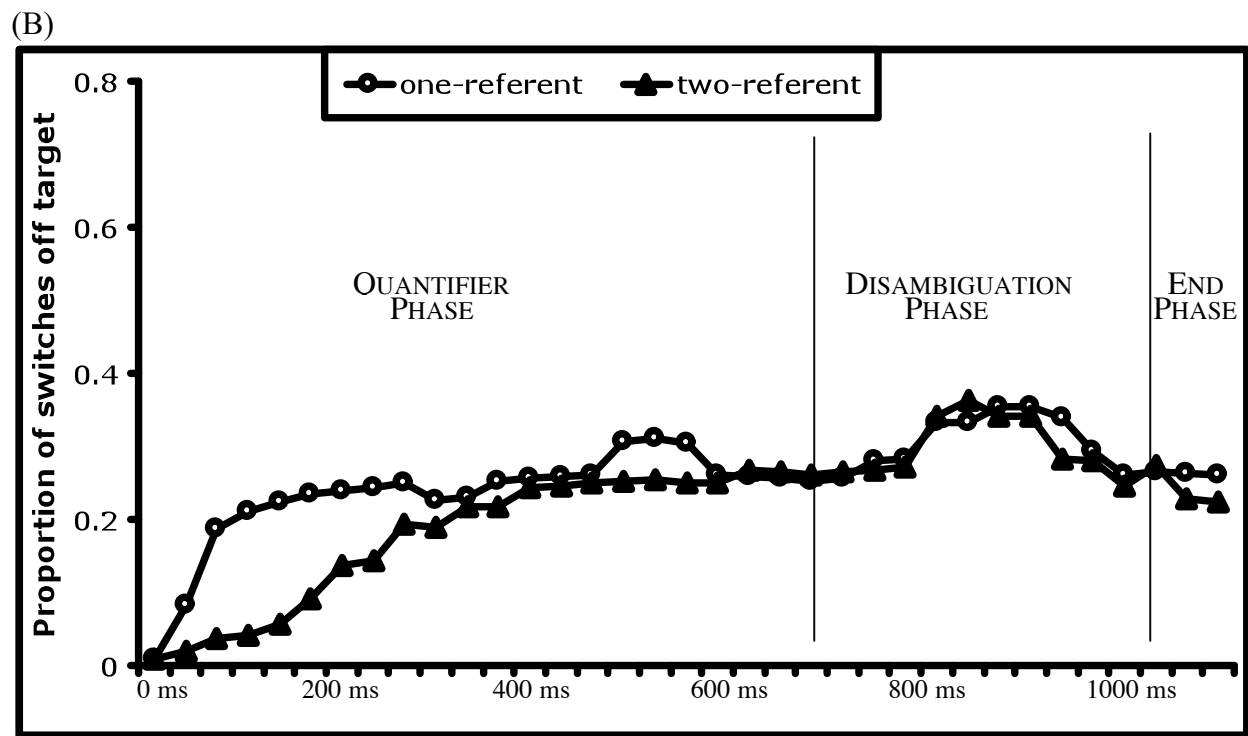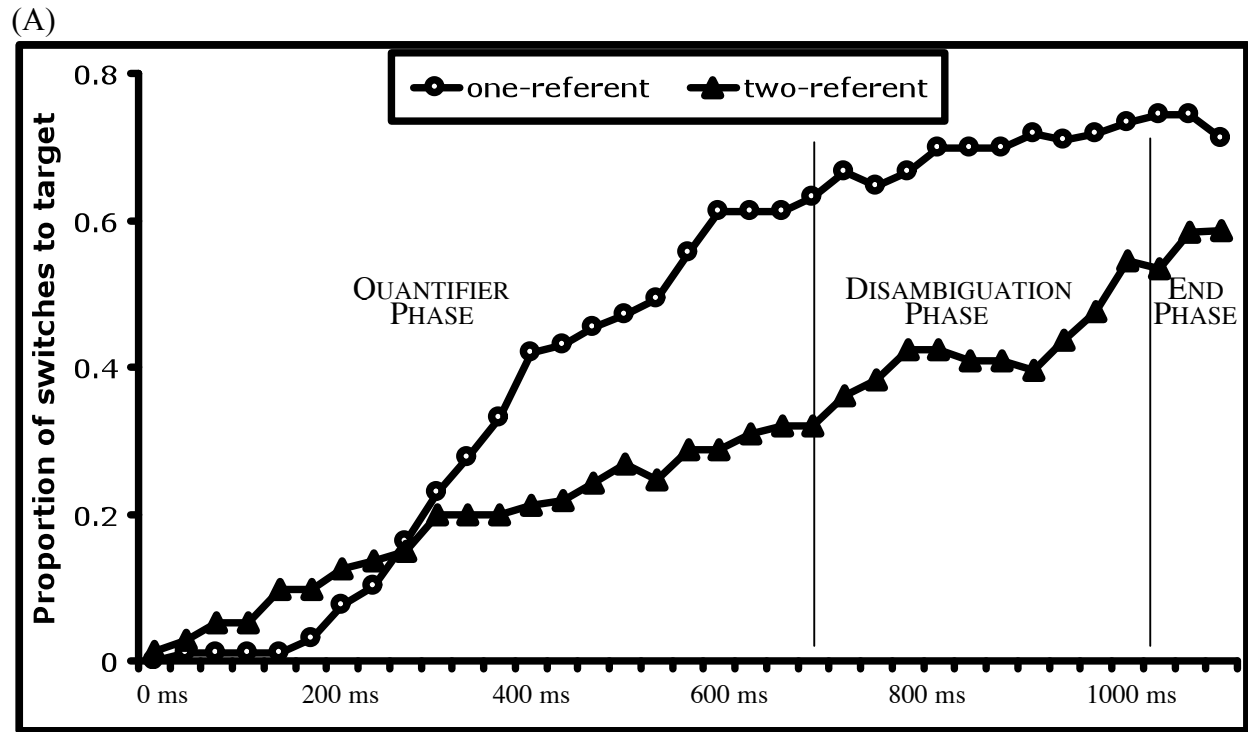
(A)

(B)
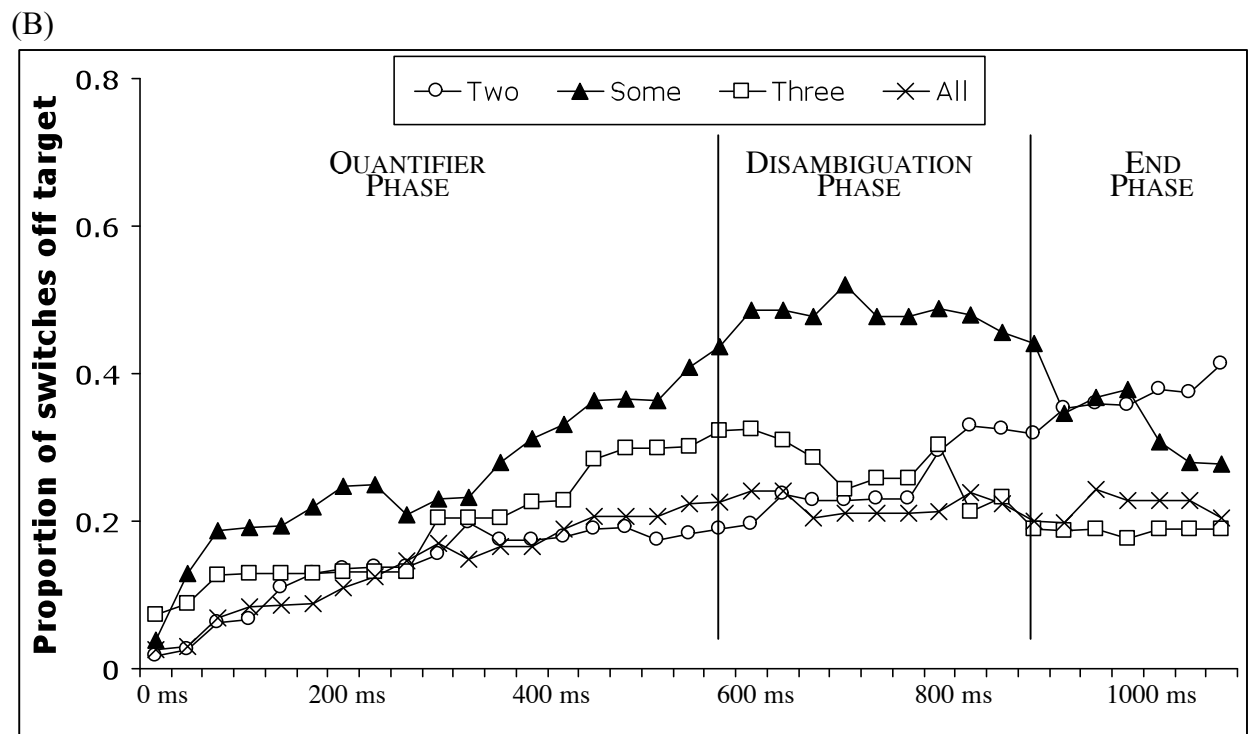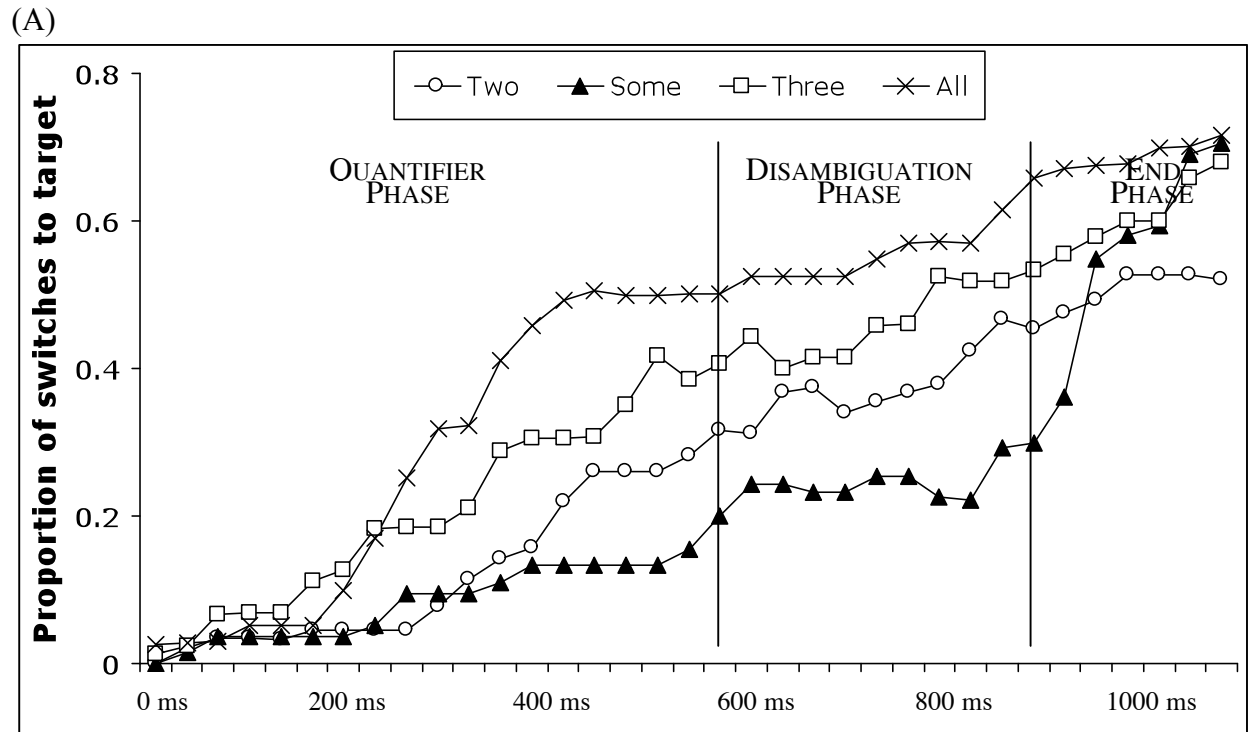
Figure 5: In Experiment 2, example of a visual-world display for the (A) 1-referent trials and (B) 2-referent trials. Participants were instructed to "Point to the girl that has some of the socks."

(A)
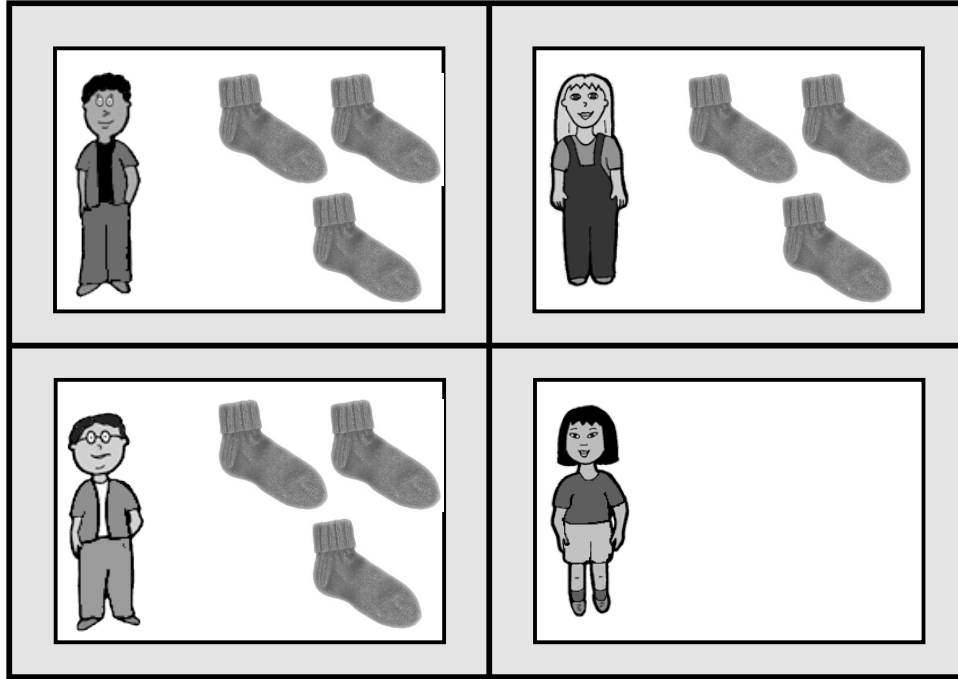
(B)
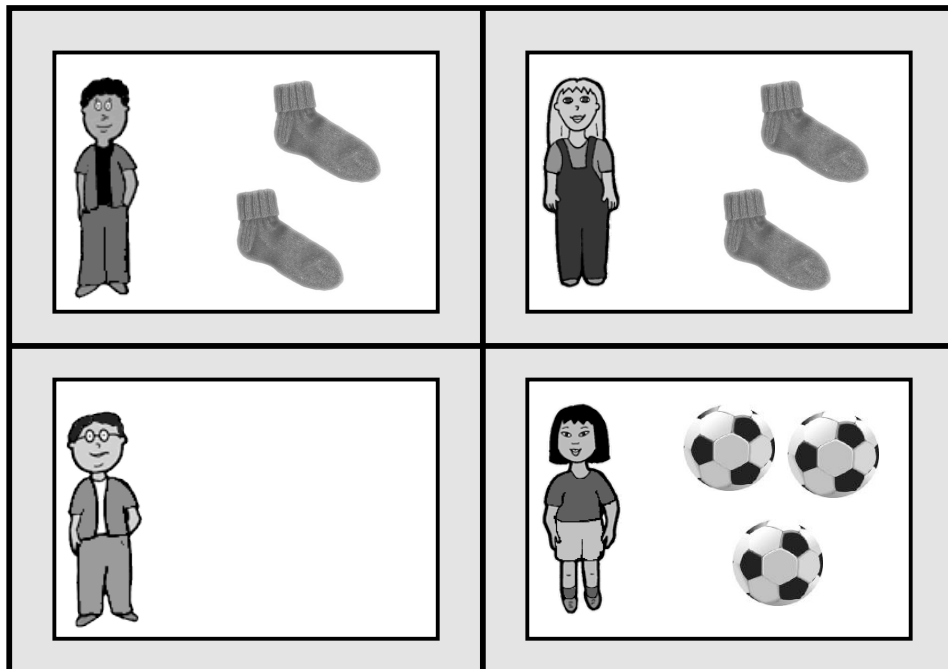
Figure 6: Experiment 2, children. Looks to the Target for the two trial types in coarse-grained time windows
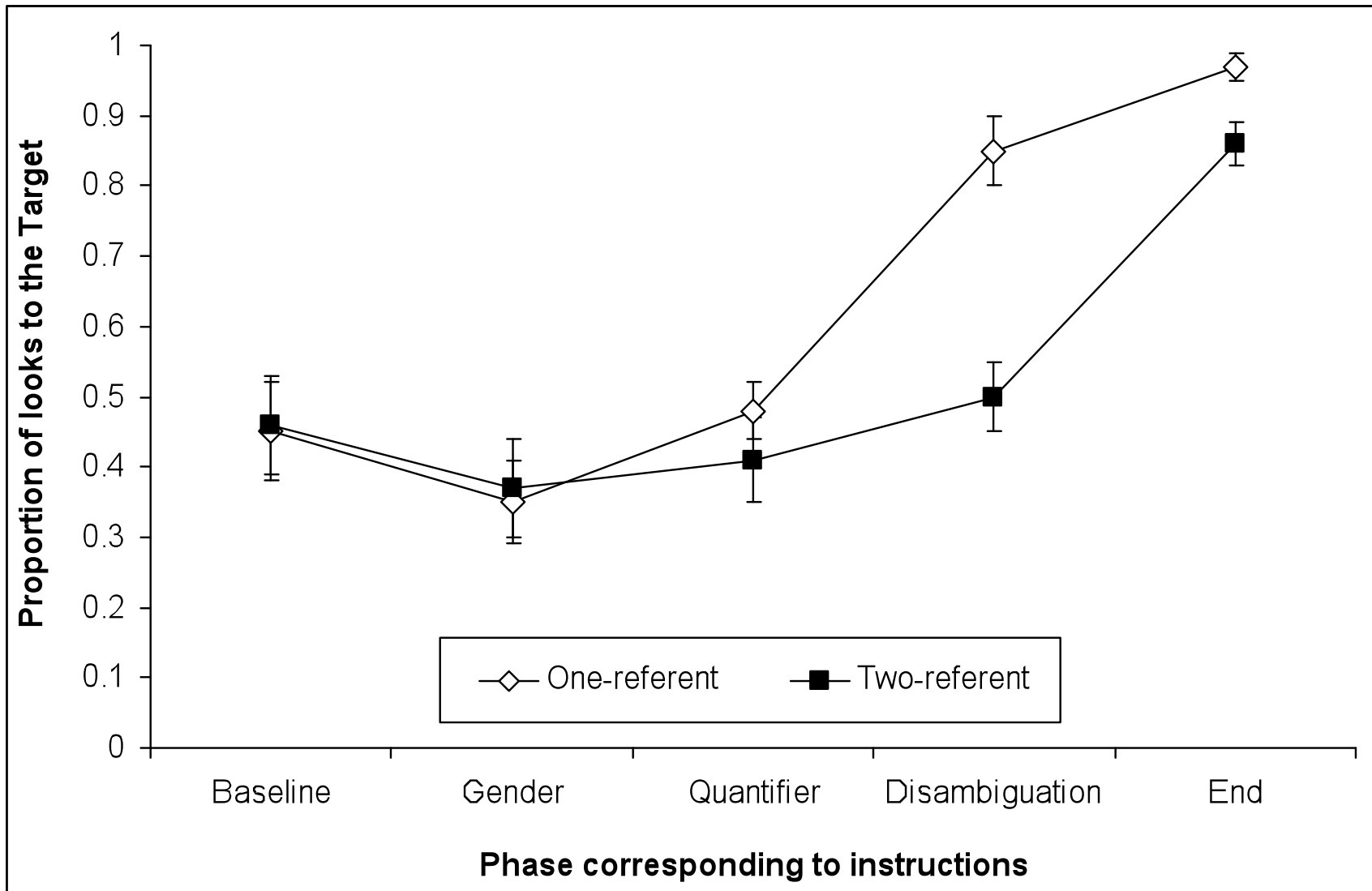
Figure 7: Experiment 2, children. Trials were separated based on fixations prior to the onset of the quantifier (A) Distractor-initial trials: proportion of switches to the Target and (B) Target-initial trials: proportion of switches off the Target

(A)

(B)

Figure 8: In Experiment 3, example of a visual-world display for the (A) SI-violating trials and (B) SI-consistent trials. Participants were instructed to "Point to the girl that has some of the socks."
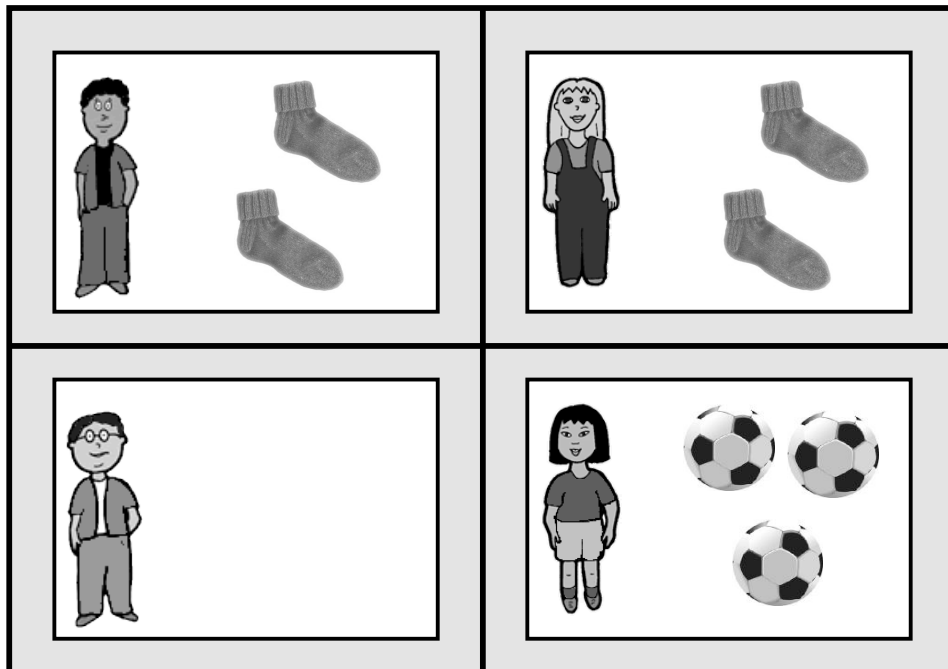
(A)
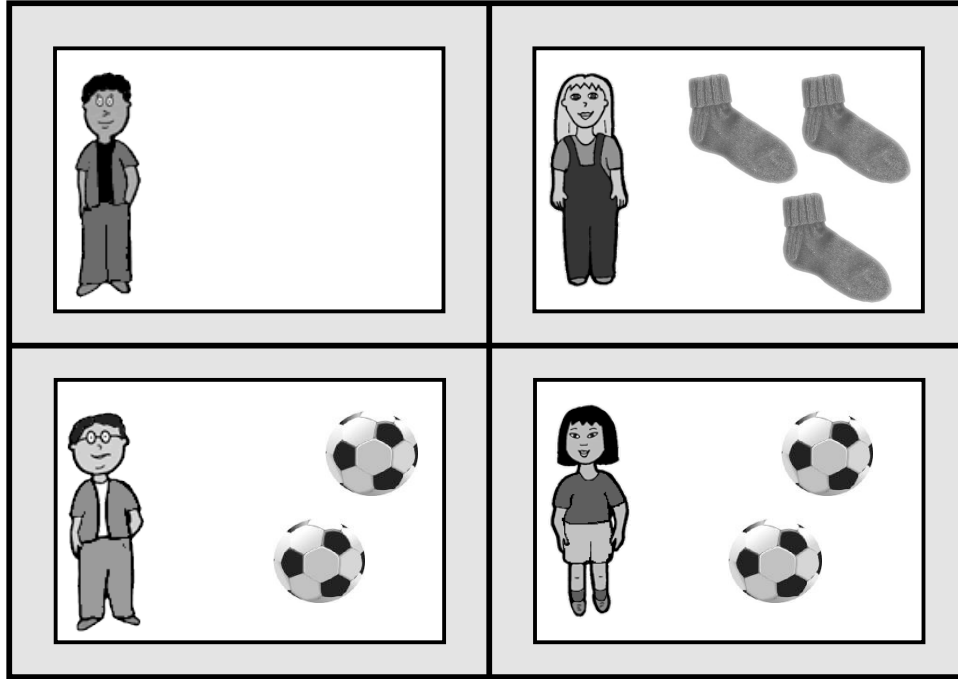
(B)

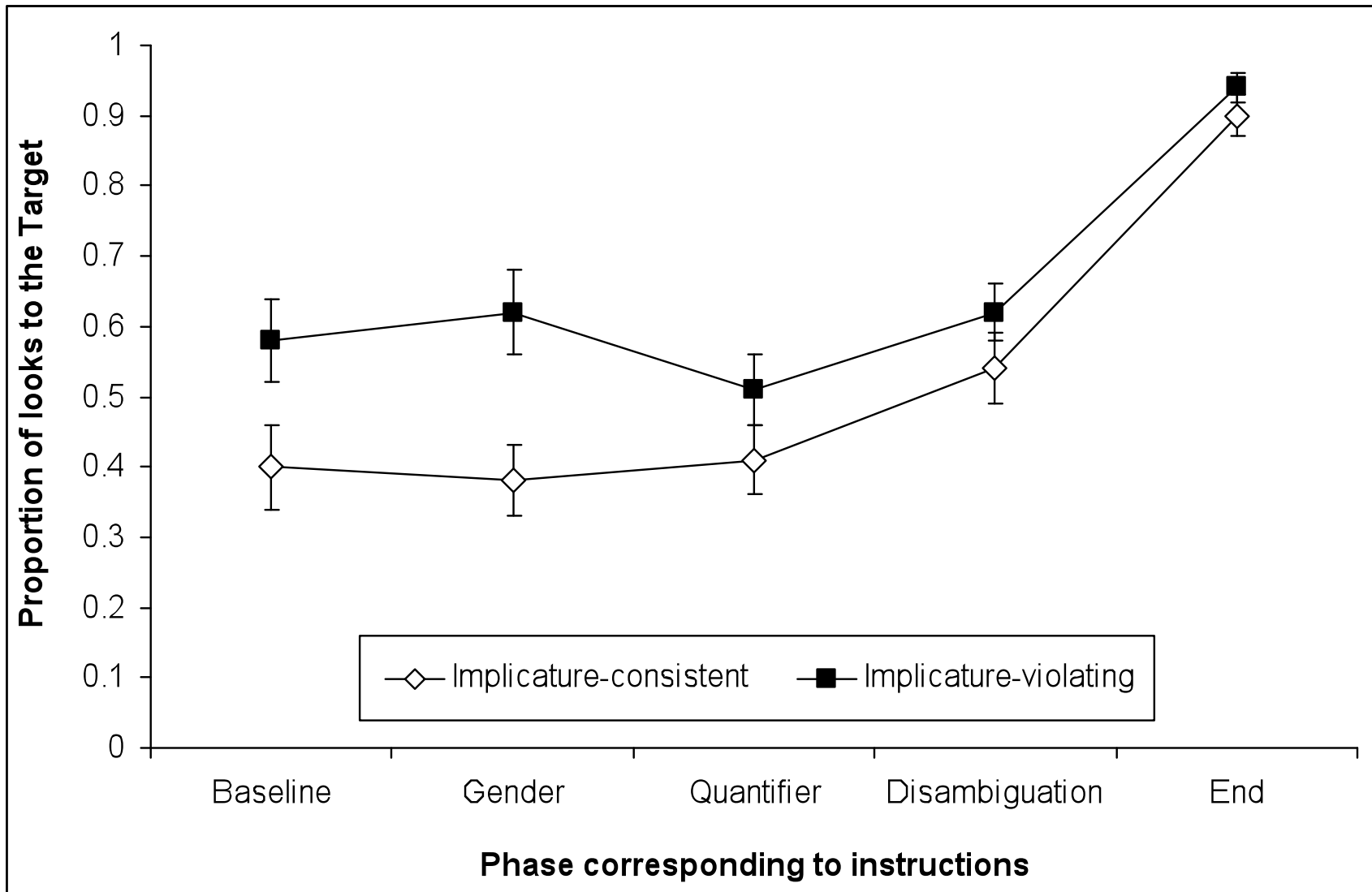Figure 9: Experiment 3, children. Looks to the Target for the four trial types in coarse-grained time windows

Figure 10: Experiment 3, children. Trials were separated based on fixations prior to the onset of disambiguation (A) Distractor-initial trials: proportion of switches to the Target and (B) Target-initial trials: proportion of switches off the Target
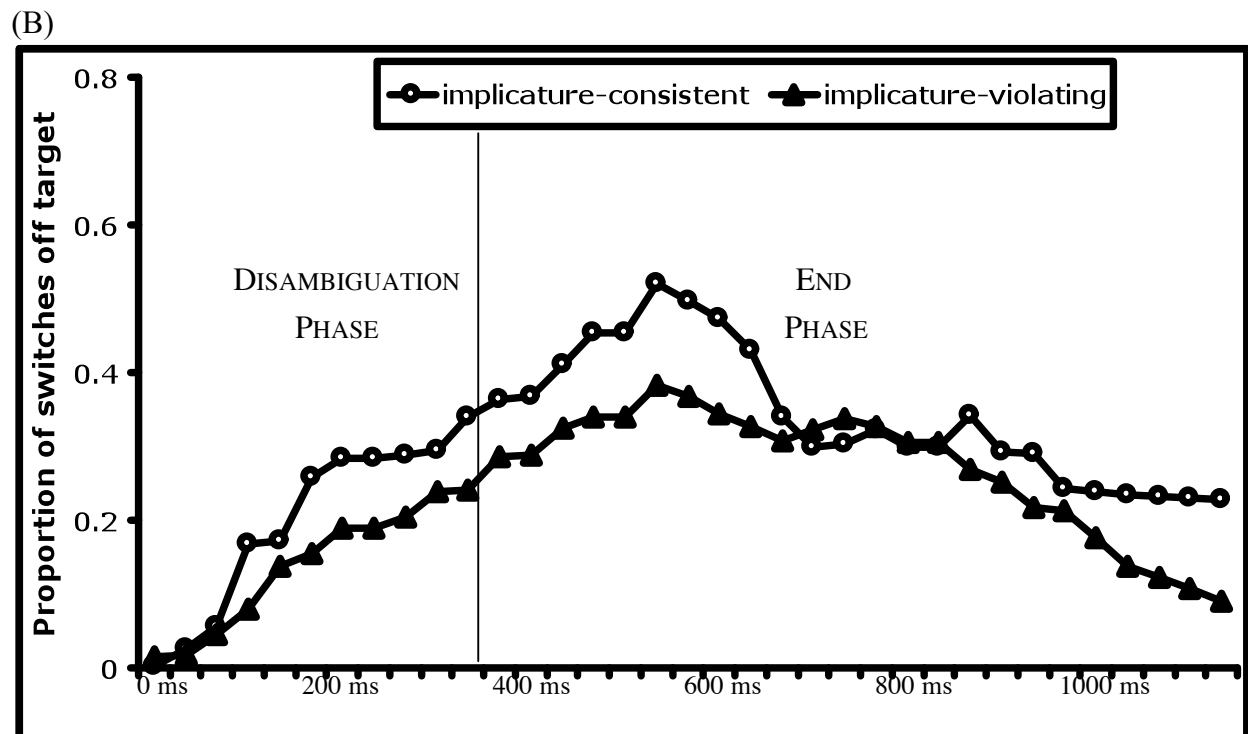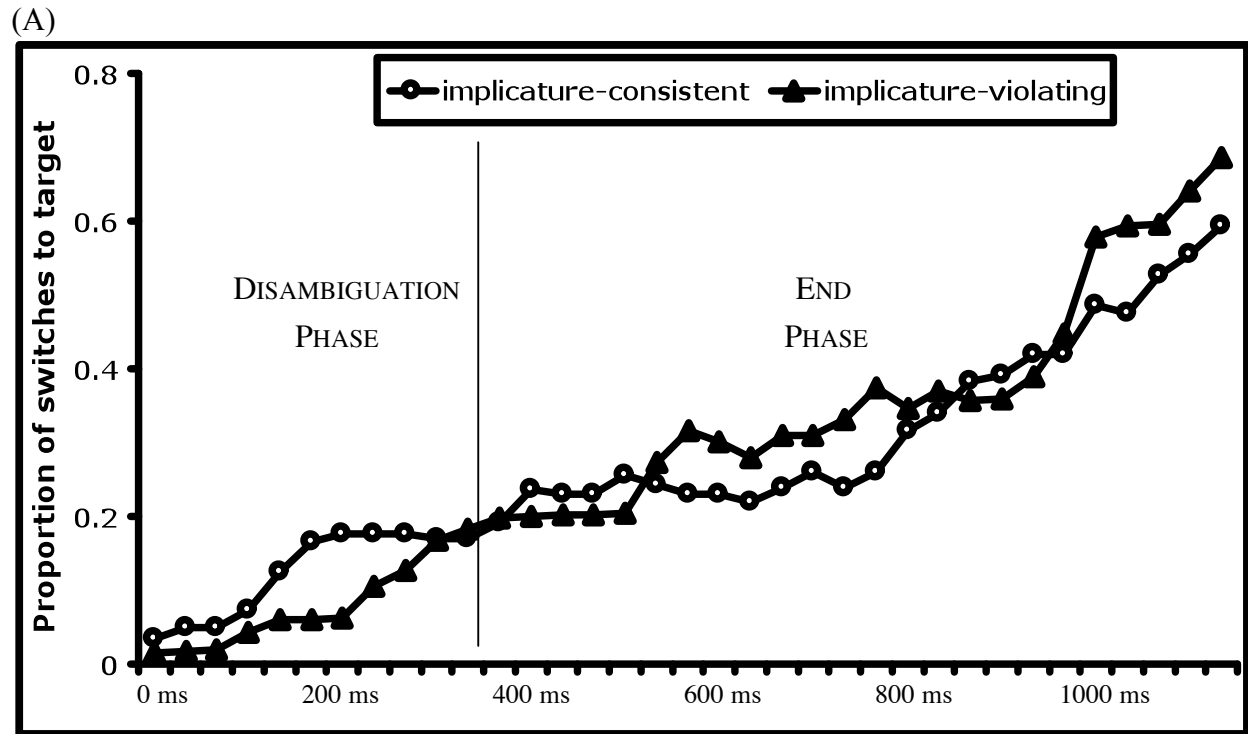
(A)

(B)

Table 1: In Experiments 1 – 3, the duration of the time windows for the course-grained analysis

	Baseline phase (POINT TO THE)	**Gender phase** (GIRL THAT HAS)	**Quantifier phase** (_____ OF THE SOC-)	**Disambiguation phase** (-KS)	**End phase** (SILENCE)
Exp 1	467 msec	533 msec	667 msec	333 msec	900 msec
Exp 2	533 msec	600 msec	733 msec	400 msec	733 msec
Exp 3	567 msec	733 msec	700 msec	467 msec	700 msec