The Development of Pragmatics: Children's Knowledge of Scalar Implicatures

Matthew Cohen

Honors Degree with Distinction Candidate

Anna Papafragou, Ph. D.

Faculty Advisor, First Reader.

1. Introduction

Implicitly we know that words only convey only half of what we want to communicate. For example, when a romantic interest says "I had a great time tonight", it makes a big difference whether she rolls her eyes or not. In fact, much of communication runs parallel to the words that are explicitly spoken and a speaker's full meaning and intent cannot be gleaned from his or her words alone. People at both ends of the conversation have to "fill in the gaps" in order to piece together incomplete statements, to excuse slips of the tongue, to guess the referents of pronouns and descriptions, and to fill in the missing steps of an argument or conversation. Without any interpolation of who *he* is, for example, the following two utterances are seemingly unrelated (from Pinker, 1994):

> Woman: I'm leaving you.
> Man: Who is he??

There is an important distinction to be made between *semantic* and *pragmatic* aspects of language. Semantics refers to the linguistically-encoded meaning explicitly expressed in words and sentences. Pragmatics refers to the principles that underlie the effective and appropriate use of language in context. What is interesting is that the two are acquired differently throughout development. The overarching goal of this paper is to explore how well children can use and understand certain pragmatic domains of language and how their pragmatic use of language develops over time.

A wealth of research has been conducted on the semantics of language acquisition and use: how children build a lexicon, how they learn and understand grammar, etc. In response to the research question "how do children use language pragmatically?," few studies can be more specific than concluding "not well." Children are notoriously bad in reasoning pragmatically with language, as is seen in their poor ability to comprehend metaphor and irony (Shatz, 1980; as cited in Papafragou and Musolino, 2003). The present investigation aims to explore children's early pragmatic reasoning in using a particular type of pragmatic inference called a scalar implicature (to be fully explained in the next section). Whereas early research concluded that children simply are not skilled at computing this kind of inference, research within the past decade has suggested that children's performance improves with different experimental paradigms. The present research explores the boundaries of children's pragmatic inferential abilities by examining their performance across new experimental paradigms.

1.1 The maxims of conversation

Language is a connection between minds and is constrained by tacit rules between interlocutors. As first noted by Paul Grice (1989), communication heavily relies on a mutual expectation of cooperation between the speaker and listener. "The speaker implicitly guarantees that the information to be conveyed in speech is relevant: that it is not already known, and that it is sufficiently connected to what the listener is thinking that he or she can make inferences to new conclusions with little mental effort. Thus listeners tacitly expect speakers to be informative, truthful, relevant, clear, unambiguous, brief, and orderly" (Pinker, 1994). Grice (1989) elucidated these implicit expectations ("maxims") of conversation. His "quantity maxim", for example, describes how much information should be included in an utterance:

i: Make your contribution as informative as is required.

ii: Do not make your contribution more informative than is required.

Grice goes on to propose a "quality maxim":

i: Do not say what you believe to be false.

ii: Do not say that for which you lack adequate evidence.

Grice's assertions delineate some of the tacit constraints that underlie pragmatic inferences. Take, for example, the use of conversational implicatures. Also called a Gricean implicature, a conversational implicature is an inference that goes beyond the explicit linguistic meaning of an utterance (Noveck, 2001). Grice believes that a listener searches for implicatures when one of the maxims is violated. In the following example (Noveck, 2001), Bettie's reply violates Grice's quantity maxim; she was not as informative as was required:

Pierre: Are the cakes ready?

Bettie: Some are.

Upon hearing Betty's response, most listeners would take this to mean that *not all* of the cakes are ready. This is not explicitly encoded (semantically) in the dialogue, however. In fact, logically, *some* means *at least some* ***and perhaps all*** (Noveck, 2001). *Some* is logically embedded in *all*. For instance, the statement ***all*** *the cakes are done* means also that ***some*** *cakes are done*. However, the statement ***some*** *cakes are done* does NOT also mean that ***all*** *cakes are done*. Thus, the statement *all cakes are done* logically entails that *Some cakes are done*. Supposed Pierre owned a bakery and was referring to a customer's order for four cakes and that his employee, Bettie, had just completed making all four. Responding "Yes, all of the cakes are done" would follow Grice's rules of informational quantity and Pierre would know that "all" means 4 out of 4 cakes are done. Because she responded "some are," Pierre only knows that *at least one* cake was completed. It's logically possible, following the meaning of the word *some*, that 4 out 4 cakes were baked. Because *some* is not very informative, a "weak" term, Pierre was forced to perform further inferences to determine whether or not Bettie was in fact done.

Because Bettie's statement was not as semantically informative as it needed to be (violating Grice's quantity maxim), Pierre was forced to use a pragmatic inference. Horn (1992) explains that when one utters a relatively weak term, e.g. *some*, it is an indication that the speaker deliberately chose not to say a stronger term, e.g. *all*. Implicitly Pierre was calculating*: if Bette had meant 4 out of 4 cakes, she would have said "all". Since she didn't say "all", it is likely that fewer than all are done*. Essentially a listener compares *what* <u>*was*</u> *said* to a set of alternative responses (i.e. what *could have been said),* thus pragmatically inferring the speaker's meaning. Consider the follow examples as well:

> Person A: How was the movie?
> Person B: It was fine.

Semantically, *fine* is a word that describes the movie favorably. Pragmatically, however, the listener compares what *was said*, "fine", to what *could have been said* (e.g. "excellent") to understand the strength of the speaker's enjoyment. Similarly, consider:

> Person A: How's your paper coming along?
> Person B: I started to write it.

Semantically, person B's statement could mean that he wrote all of the paper. Person A, however, is implicitly comparing what he did say, *started*, to what he could have said, *I finished writing it*, to infer that B did not finish the paper. Because the listener is inferring a scale when making such inferences, terms that fall along a scale are called **scalar items** (SIs), and the inference that the stronger term on the scale does not hold is called a **scalar implicature**.

In short, the issue is that, logically, the weaker SI (e.g. some) doesn't have to mean not-the-stronger-SI (e.g. *not all)* even though it is used to mean *not all* pragmatically. Linguists would say that, as a quantifier, the lower bound of "some" is determined <u>semantically</u> to mean *at least 1*, and the upper bound of "some" is determined <u>pragmatically</u> to mean *not-all*.

Because the pragmatic use of *some* is so ubiquitous among adults, this fine distinction between semantics and pragmatics hardly seems important. What is interesting, however, is that children do not place a pragmatic upper bound on *some,* as adults do, as we go on to see in the next section.

1.2 Development of Pragmatic Inference

Children are as comfortable accepting the statement *Some birds have wings*, as well as the statement *All birds have wings* because their use of *some* is governed only by its semantic lower bound. A study using sentences like these was first done by Smith (1980), using 4-7 year-old children. Children in her study were asked to respond yes or no to questions like *Do all elephants have trunks*? Smith found that 4-7 children accept statements that overextend the pragmatic upper bound of *some*, e.g. they accept the statement *Some elephants have trunks* when *all* would have been a more informative, pragmatic choice. Although Smith did not include adult subjects, she reasoned that adults would reject the statement (in favor of a stronger SI, like *all*). Noveck (2001) replicated (with slight modifications) Smith's (1980) study but included subjects at three ages: 7, 10, and adult subjects. Results showed that at age 7, 84% of children correctly responded to sentences that appropriately used *some* (e.g. *Some birds live in cages*), 91% of children correctly responded to sentences using *all* (e.g. *All elephants have trunks*), and 89% of 7-year-olds accepted true but pragmatically infelicitous statements (e.g. *Some giraffes have long necks*). 85% of 10 year olds and 40% of adults accepted pragmatically infelicitous statements. Noveck's results show that over time people begin pragmatically separating the domains of *some* versus *all*. Noveck (2001) also found similar results with the same age groups for statements with ***might*** *be X* when the context indicates that the stronger ***must*** *be X* would be a more informative response.

Papafragou and Musolino (2003) asked children and adults to evaluate true but infelicitous statements using *some/all*, *start/finish*, and the numerals *two/three* (numerals also fall on a scale). Instead of asking children to agree or disagree with the statements (as Noveck 2001 did), the experimenters asked whether or not the speaker "answered well." Modifying the question in this way highlights the felicity of the target statement and not the truth, since underinformative statements are still technically true. Like Noveck (2001) their results showed that children consistently interpret scalar expressions logically, without applying a pragmatic upper bound. 5-year-old children rejected only 12.5% of the infelicitous statements using *some*, 10% of the infelicitous statements using *start*, and 65% of the infelicitous statement using numerals (the special domain of numbers will be discussed later). Adults rejected the infelicitous statements 90+% of the time for all conditions. In a follow up experiment, Papafragou and Musolino (2003) assessed the influence of experimental conditions on children's performance. They included a pre-test training to prime children's detection of pragmatic anomaly and provided more specific instruction to highlight the goals of the task. The speaking character in the experiment, "Minnie," was described to sometimes "say silly things," in which case the child should correct her to help her "say things better." After the training and with the modified design, children performed much better: they corrected 52.5% of pragmatically infelicitous statements using *some*, 47.5% of those statements using *start*, and 90% of the infelicitous number statements.

The importance of experimental conditions was further addressed in Papafragou and Tantalou (2004), where 5-year-old children were instructed to award prizes to animal characters if the characters had completed assigned tasks. Depending on the animals' responses (using weak or strong SIs), children had to decide whether or not to award them the prize for completion. For example, if the experimenter asked the animal character, "Did you color the stars?" and the animal responded "I colored some," a child pragmatically inferring *not all* would not award the animal a prize. Children in this experiment refused to give a prize (i.e. pragmatically inferred non-completion) 77.5% of the time.

Although no experiment shows children's understanding of scalar implicatures to be at an adult level, their relative performance can be very different depending on the nature of the experiment. Smith (1980) and Noveck (2001) seem to under-represent children's abilities. One reason might be that the quantifiers (*some, all*) that they used were out of conversational context and did not evoke a strong pragmatic inference of what was intended. A second problem might arise from the children's response format. The studies asked children to say *yes/no* (Smith, 1980) or to agree/disagree (Noveck 2001) with the statements. Responses like these tap the children's assessment of the *truth* of the statement (relying on semantics), rather than the *felicity* of the statement (pragmatics). On the other hand, a problem with studies that ask children to correct the "silly" puppet might be that children may not compute an implicature from the conversation because they might simply dismiss the puppet as "silly" and not even try to make pragmatic inferences from the statements of such a speaker.

1.3 Experimental Plan

The goal of the present research is to explore children's understanding of pragmatic inference, particularly the factors and conditions that might improve children's performance with scalar implicature. The current studies modified ways of assessing children's understanding of SIs. Study 1 asks children to listen to a story and produce the end result of actions that are modified by a scalar term. Essentially the study asks children to turn a SI into a number.

Previous work has asked children to assess or correct the statements of others. This study is interested in how children themselves spontaneously interpret scalar terms.

In order to fully understand the motivation for studies 2 and 3, recall Grice's maxim regarding interlocutors' expectation of quality:

i: Do not say what you believe to be false.

ii: Do not say that for which you lack adequate evidence.

If a speaker is fully informed of an event, he or she is in a position to use strong terms to describe it. Conversely, a speaker who is not adequately informed of an event will most likely use a weaker term since their use of a stronger term would violate Grice's second submaxim of quality. In order for children to succeed in previous studies (e.g. Noveck, 2001), children needed to implicity understand Grice's first submaxim of quantity (i.e. Make your contribution as informative as is required) as well as the second submaxim of quality (Do not say that for which you lack adequate evidence). Essentially, in order to compute a scalar implicature, one needs to implicitly understand that an utterance contains as much true information as the speaker is capable of conveying. Studies 2 and 3 in the present investigation further explore the interaction of the two maxims. Specifically, they make Grice's quality maxim more salient by manipulating two potential speakers' access to evidence. In these experiments, two characters witness an unfolding event. One speaker has access to the full event, while the other only has partial access (as a result of "falling asleep" halfway through). At the end of the event (and without any evidence of the end product), the child hears a description of the event that perceptually could have come from either speaker. Based on the content (i.e scalar terms) of the utterance, the child is asked to determine which of the two characters was the speaker. Correct responses in these studies pair event descriptions with weak scalar items to characters with partial knowledge and strong scalar items to characters with full knowledge. Using Grice's terminology, we are interested in seeing whether or not children can match statements with weak informational terms (e.g. *some*) with speakers who "lack adequate evidence", and also matching informationally strong terms (e.g. *all*) with speakers who do have adequate evidence.

Structuring the experimental questions in this way targets abilities that children are proficient at when they are 5 years of age. Numerous studies have shown that children are in fact quite good at reasoning about properties of speakers. For instance, studies have shown that children are able to monitor a speaker's reliability (Koenig et. al., 2004), and take account of a speaker's knowledge when learning a novel word from them (Sabbagh & Baldwin, 2001). The question we are now asking is whether children can use information about a speaker's knowledge to attribute an utterance to that speaker, in accordance with Grice's maxims.

The studies that follow use different scalar terms and methods to explore these issues. Study 1 asks children to physically produce the results of short narrated stories with scalar terms. They are not asked to clarify or assess the comments of others, but rather to translate scalar terms (<*some,all*> and <*start,finish*>) into a discrete quantity of tangible things. For example, one task asked the child to put *some of the apples* onto a plate when there are 4 apples present. The pragmatically sophisticated answer would be to place one, two, or three apples on the plate, whereas an answer bound only by the lower semantics of *some* might place all 4 apples on the plate. Study 2 asked children to attribute statements with **quantifiers** like *some* and *all* to different potential speakers and aim to determine whether children apply quantifiers using the *at least* semantics (even when it is underinformative to do so) or whether they can place a pragmatic upper bound on quantifiers by computing a scalar implicature. Study 2 also asks children to attribute statements using **numbers** instead of quantifiers. We are interested in comparing children's attribution of statements using numbers because, unlike quantifiers, numbers are commonly used with an *exact* semantics (as opposed to an *at least* semantics). Finally, Study 3 uses a similar logic as Study 2, but instead of using terms that fall along a natural, logical scale, like quantifiers and numbers, children are asked to attribute statements to speakers who have different knowledge of items on an ad hoc scale (Horn, 2005). We use 4 **transformation events** to explore this area, where one object is drawn on a piece of paper and is then transformed into a second object. Gricean principles would dictate than speakers with knowledge of the complete drawing would describe their full knowledge and not report their knowledge of the first drawn object only.

2. Study 1: Spontaneous Understanding of Scalar Items

The first study asks 5-year-old children to listen to a story and reproduce certain actions described by a scalar quantifier/verb. For example, children are asked to show how Freddie Fish (hand puppet) performed actions like "putting some of the apples on the plate" when there are four apples present in the scene. The scalar terms are *some*, *all*, *start*, and *finish*. One advantage of this task is that it doesn't ask the participants to use any judgment of truth or felicity – all they have to do act out their (lower or upper bounded) interpretations of a scalar term. The purpose of this experiment is to uncover how children spontaneously understand scalar terms.

2.1 Participants

Participants in this experiment were 16 English-speaking children between the ages of 4;5 and 5;9 (mean age 5;1) recruited from kindergartens in Newark, DE. Six were male. Their parents signed consent forms approved by the University of Delaware Institutional Review Board.

2.2 Materials and Methods

Children were removed from their classrooms for about 10 minutes with their assent and tested in a quiet room away from their peers. On a small table, there were objects placed in various locations, creating tasks for the children to perform. Some of these tasks involved multiple iterations, like pretending to water (with a toy watering can) a series of 4 potted (plastic) flowers. Others were "filler" actions, designed to mask the purpose of the experiment. Filler actions only required one iteration, like putting a single toy bird in a small wooden birdhouse. Verbal instruction on how to complete the filler task did not require modification by a scalar item. The experimenter had a fish puppet, "Freddy Fish," and children were told that they would be hearing a story about things that Freddie Fish did (i.e. the tasks) and that they had to replicate what Freddie did (i.e. complete the task).

> *I'm going to tell you a story about some things that*
> *my friend Freddie fish* [indicate hand puppet] *did.*
> *Do you think you can show me how he did them?*

Children were randomly assigned to one of two conditions. Children assigned to the Quantifier condition were told that Freddie Fish did ***some*** of the (four) possible actions on the test (i.e. non-filler) tasks. Children in the Verb condition were told that Freddie ***started*** doing the (four) possible actions in the test tasks. For both conditions, children saw 4 test and 4 filler trials in a pseudo-random order. Both conditions concluded with two strong SI tasks (with *all* or *finish*). The strong SIs were added to the end of the task as control trials, so that children would not use the contrast between *some* and *all* and be aided in their interpretation of *some* as *not-all*. Correct answers for the weak SI trials (*some* and *start*) were one, two, or three tasks completed out of the possible four. The only correct response for the strong SIs (*all* and *finish*) were four out of four tasks completed.

[INSERT TABLE 1 HERE]

Within each condition the order of the trials (test items and fillers) was counterbalanced between children. Two strong SI trials (control) were kept at the end but their internal order was also counterbalanced.

2.3 Results

Results from a one-way ANOVA with the number of upper-bounded responses (1,2,or 3 actions) as the dependent variable and Condition (Quantifier, Verb) as a between-subjects variable yielded no significant effects of Condition: $F(1,15)=.4986$, $p=.49$.

[INSERT FIGURE 1 HERE]

In the Quantifier condition, children's mean responses show that the group was evently split between taking *some* to mean 1,2,or 3 out of 4 items (50% of responses) versus all 4 items (50% of responses. In the verb condition, children interpreted *start* infelicitously (4 out of 4 items) 65% of the time. While the group means from this study seem to indicate that all children perform at chance, the numbers actually reflect 2 distinct groups of participants: those who pass the majority of their test items and those who consistently fail them.

[INSERT TABLE 2 HERE]

Table 2 reflects the number of participants that fall neatly into a "passing" or "failing" category (defined as performing consistently either way on 3 out of the 4 test items). With the exception of 2 (out of 16) participants, all children fell neatly into a passing or failing category.

We also looked at the exact number of actions (out of 4) that children tended to prefer for the Quantifier and Verb conditions. In the Quantifier condition (*some*), children tended to produce 1,2 or 3 actions equally, whereas in the Verb condition (*start*) they were more biased towards performing only 1 action in the test items.

[INSERT TABLE 3 HERE]

ANOVA with the number of exact (4/4) responses on the strong SI (control) trials as a dependent variable and Condition as a between-subjects variable did not yield significant results: $F(1,15)=1$, $p=.33$, All participants performed correctly on both filler items.

2.4 Discussion

Depending on children's understanding of SIs, the study could have yielded two possible results with *some,all*. If the children were to make pragmatic inferences based on the strength of the scalar term (*some* apples vs. *all* apples), they would have performed between one and three actions on the test items (e.g. putting between one and three apples on the plate), consistent with the semantic lower bound of *some* and the pragmatic upper bound (*not-all*). However, if they were to only consider the semantic lower-bound, as previous studies have indicated that they tend to do, they should have performed anywhere between 1 and 4 iterations of the task. The most interesting analysis is whether or not they did all 4, the semantically true-but-infelicitous response.

At first glance, our results seem consistant with the second explanation (and with the literature). It initially appears that children only consider the semantic lower bound of *some* and *start* and are confortable extending the words into pragmatically-infelicitous situations. While the groups appear to be split, closer analysis reveals that each child is remarkably consistent. One group of participants reliably places a pragmatic upper bound on the weak scalar items and perform 1 to 3 actions (out of 4). A second group of our participants reliably overextended the weak scalar term. We are unsure how this finding aligns with the literature since most studies only publish group means. If this is a reliable finding between studies, it is an important point to highlight since it implies that at age 5, children's pragmatic abilities are sensitive to individual differences, not ubiquitous developmental milestones.

While there are two groups of differing pragmatic abilities ("passers" and "failers"), there are more "passing" children in the Quantifier condition than in the Verb condition. As noted by Papafragou and Musolino (2003), scalar items are not necessarily a homogeneous group. For a discussion of why *start* might have different properties from other SIs (e.g. *some*), I defer to Papafragou (2006, p.726):

> The acquisition of the verb *start* "requires grasp of some vexing aspects of scalar logic. Even though the verb semantically denotes the initial part of an eventuality, it is not clear how that part is specified. When is it possible to say

> *John has started to bake a cake?* After he has consulted the recipe? After all the ingredients have been assembled on the counter? After the packet of sugar is opened? After the mixing has begun? The answer seems to depend on complex consideration, including intentionality, expectations about the normal course of events, etc…"

Thus, it is possible that children's performance in the Verb condition (versus participants in the Quantifier condition) was influenced by the difficult nature of what it actually means for Freddie Fish to *start planting the flowers*. It is also interesting to note that children's actual responses to test items with *start* were different than their responses to *some*. Recall from Table 3 that children were equally comfortable performing 1,2, or 3 actions to denote *some*, whereas the majority of participants performed only 1 action to show that Freddie Fish had *started* to do something. This indicates that *some* and *start* possess different pragmatic properties for the child, where pragmatic uses of *start* tend to mean dramatic non-completion (something like *barely-started)*.

The overall conclusion from this study is that children are not very successful at spontaneously computing scalar implicatures. One limitation of this method, as with some previous studies (e.g. Noveck, 2001), is that the pragmatic inference paradigm is not based in interaction (e.g. dialogue) so perhaps potential pragmatic inferences would be neglected. Studies 2 and 3 ask children to attribute a statement to speakers who are describing an event to a third party. It is our hope that this paradigm will draw more heavily on children's abilities to pragmatically infer information.

3. Study 2: Speaker Knowledge and Scalar Terms

In this study, children (and adults) were asked to match statements to differently informed speakers. All participants saw the same videotaped stories where 2 puppets gained differential access to an event. One group of participants heard statements use either the informationally-weaker quantity *some* or the informationally-stronger quantity *all*. *Some* is a "weaker" term because semantically it doesn't specify the exact quantity it describes (it might be 25% of the whole or 100% of the whole). We anticipated that our participants will ascribe statements with *some* to speakers who are only partially informed, (implicitly reasoning that the speaker would have used in a stronger term (i.e. *all*) if they were in a position to do so), and would reserve statements with *all* for fully-informed speakers.

The second group of participants saw the same videos as the participants who heard statements with quantifiers, but instead heard statements that described the events with the numbers *two* or *four*. As Papafragou and Musolino (2003) note, not all scalar orderings can be treated the same way. Numerals, for example, tend to be assigned an *exact* interpretation, rather than an "at least" semantics, as is the case with *some* (Hurewitz et. al, 2006). A three-sided figure is not a figure with *at least* three sides, it has *exactly* three sides. Also consider the following exchange:

> Person A: Can you pass me three napkins?
> Person B: I can pass you two.

In considering the preceding example, it is unlikely that someone would treat the number 2 with only a semantic lower bound and take "two" to mean *at least two, maybe more*. Unlike *some* and *start*, numerals are not upper-bound by a scalar implicature and any scalar inferences drawn from conversation with a numeral would not be considered *implicatures*; one can succeed in this task by following Grice's maxims of saying what you know to be true to the extent that is sufficiently informative. We are interested to see how the knowledge-state paradigm would transfer to scalars with *exact* semantics. Specifically, we expected that children will successfully attribute statements using *two* or *four* to speakers with different knowledge better than statements with *some* or *all*, since numbers are taken to refer to exact quantities (rather than a lower boundary) and thus less cognitively demanding than attributing statements with *some/all*.

3.1 Participants

28 children and 16 adults participated. Children's ages ranged between 4;4 and 5;11 (mean age=5;0) and the adults ranged between the ages of 18;8 and 21;8 (mean age = 19;7). Child participants were recruited from Newark, DE, and had parental consent. The 16 adult participants were selected randomly from a pool of students enrolled in General Psychology. All adults were 18 years or older and gave their informed consent. All participants were different from the participants in Study 1.

3.2 Materials and Methods

Children were removed from their classroom with their assent and brought to a quiet room away from their peers. The children were seated in front of a laptop and the experimenter sat next to the child and operated the laptop. Subjects were randomly assigned to one of two conditions (Quantifier, Verb).

Each child was first shown two "warmup" videos so that they would feel more conformable with the types of questions they would be asked on the target videos. The warmup trials consisted of two very short (approximately 15 seconds each) videos of two animal puppets performing different actions. For example, one animal was reading a book and the other was watering a plant. At the end of 10 seconds, the video would freeze and the children would hear a voice (computer file) that would either say "*I watered a plant*" or "*I read a book*." The children were then asked "*Which animal said that? Point to the animal that said that*." All subjects correctly attributed the statements and passed the warmup trials.

The children were next shown a series of 4 short test videos. Before each video began, there would be a 10 second freeze-frame where the experimenter would introduce a human character in the video, Sarah, and 2 different animal puppets, all sitting in front of a table. The person controlling the puppets was under the table and was not visible to the child watching the video. After the characters were introduced, the video began. In each video, Sarah performed 4 iterations of an action, like planting a flower in each of 4 flower pots, while the two animal puppets "watched." After performing two of the actions, Sarah would stop and make it seem like she wasn't going to continue. At this point, one the animal puppets would "fall asleep" by turning away from the action and lying down. Sarah would then place a blanket over the sleeping animal. She would then gesture to the other animal that the first animal was asleep and that she would quietly continue. After completing the remaining two iterations (e.g. planting the remaining flowers), Sarah covered up the task so the end result wasn't visually apparent and then woke up the sleeping animal. A second human character would then walk into the scene and ask the puppets, "*What did Sarah do today*?" The video would then freeze. Subjects in the Quantifier condition heard either "***Sarah planted <u>some</u> of the flowers***" or "***Sarah planted <u>all</u> of the flowers***" (sound file from the computer). Subjects in the Number condition heard "***Sarah planted <u>two</u> of the flowers***" or "***Sarah planted <u>four</u> of the flowers***." The experimenter would then ask the child "*Which animal said that? Point to the animal who said that*". A correct response would pair the *some* or *two* statement to the animal who only witnessed half of the event, and the *all* or *four* statement to the animal that witnessed the whole event.

There was only one voice statement at the end of each video, either a "some" statement, or an "all" statement (or "two", "four"). Each child saw 2 *some*-videos and 2 *all*-videos (or 2 *two* videos and 2 *four* videos) in blocks, order counterbalanced. Additionally, to ensure that certain tasks, like planting flowers, weren't more appropriate for "some" or "all" statements, there was a second version of the stimuli, given to half of the participants, that matched the "some" or "all" statements to the opposite videos from the first version.

Each video used a different set of 2 animal puppets, to ensure that the participants wouldn't remember one animal as consistently being the uninformed (or "sleepy") one. Furthermore, the animals were placed on the same side as one another, and to Sarah's right or left (counterbalanced). The relative position of the sleepy character (nearer/farther from Sarah) was also counterbalanced. Children's ages in the Quantifier conditions ranged from 4;4 to 5;8 (mean=4;11), and children in the Verb condition ranged from 4;4 to 5;11 (mean=5;0).

Adult subjects were run in groups of 4. Each group of four sat in a small presentation room facing forward towards a projection screen that showed the same stimuli as the children. The adults did not attribute the statement verbally or by pointing, as the children did, but rather by silently circling a picture of the correct animal on their own response sheet (see Appendix 1).

3.3 Results

A two-way ANOVA with Condition (Quantifier, Number) as a between-subject factor and Statement type (Strong,Weak) as a within-subject factor and the number of correct responses that children produced as a the dependant measure yielded no main effects or interactions.

Collapsing across both the Quantifier and Number conditions, the weak statements were not different from chance, but the strong statements were ($F(1,27)=2.2$, $p=.03$).

One interesting result was that when the weak terms were presented first, children were significantly better at attributing them ($F(1,27)=7.83$, $p=.009$). The mean number correct (out of 2) for strong terms presented first was .78, where as the mean number correct (of 2) was 1.5 for weak terms presented first. This is interesting, too, when combined with an analysis of "passers" and "failers". 10 (out of 14) in the Quantifier condition were strongly passing or failing. All "passing" children (4) saw the weak statements first whereas all "failing" children (6) saw the strong statements first. "Passing" or "failing" does not seem to reflect attributes of the children themselves, but rather the condition they were placed in. Rather, children who saw weak terms first were more likely to pass them. A similar (though not as strong) relationship was seen in the Number condition. 6 children (of 14) passed (4 of which saw the weak statements first) and 2 kids failed (both saw the strong statement first).

In the Quantifier condition, our results indicated the same trends as previous studies. Half of the time, children correctly attributed the statement with *some* (the weaker SI) to the speaker with partial knowledge and 66.7% of the time extended the *some* statement to the speaker with complete knowledge (true but not pragmatically felicitous). Adults correctly attributed the weak SI statement 93.75% of the time and were very successful (100%) with the strong SI (*all*) statements.

In the Number condition, children correctly attributed the statement using *two* to the speaker who witnessed two actions 64.3% of the time, and they correctly attributed the statement using *four* to the speaker who witnessed four actions 67.85% of the time. The adult subjects, as expected, performed very well. One adult was excluded in the final analysis because she did not appear to understand the experimenter's instructions and performed below chance. With this one exception, adults correctly identified statements using *two* 92.86 % of the time and statements using *four* 100% of the time.

[INSERT FIGURES 2 AND 3 HERE]

3.4 Discussion

Our results show that children were able to correctly assign weak SI statements with quantifiers 50% of the time, presumably based on their pragmatic inferences concerning which speaker was in a position to use which quantifier. To put this number in context, recall that Noveck (2001) found that children rejected true-but-infelicitous statements 11% of the time, Papafragou and Musolino (2003) found that (without training), children identified assertions that were not "said well" 12.5% of the time (with training = 52.5%), and Papafragou and Tantalou (2004) found that children would not award prizes to animals who modified their statements of completion with a weak quantifier 77.5% of the time. Our results suggest that the findings from previous studies using violations primarily of Grice's *quantity* maxim also generalize to situations where children need to rely more heavily on quality when attributing a statement to the appropriate speaker.

We have seen that children tend to perform relatively poorly with weak SIs (e.g. *some*) and very well with strong SIs (e.g. *all*). In the case of numerals, we see that they perform in the middle, both in attributing statements with *two* and *four*. There are several points to discuss. First, children's performance on numerals compared to their performance on *some,all*, and second, children's relative performance within numerals. First, judging from the data, attributing statements with numerals seems to be easier than statements with *some*, but harder than attributing statements with *all*. We have already discussed why *some* is hard, since semantically it can mean different quantities. Perhaps *all* is easier than numerals because it is less demanding cognitively. Whereas for numerals, children would have to keep in working memory the exact amount (two or four) of iterations that each animal saw, the word *all* consistently (across situations) means something whole/complete so the children would only have to remember the relative information available to the animals, not the exact amount.

Regarding the second issue, children's performance on *two* relative to *four*, compared to *some* relative to *all* might be explained using the same logic. Even though *four* represents full information, it still is an exact amount, not an upper bound. *Four* is not easy compared to *two* because they are both numerals that represent exact quantities. *All* is simply the maximum and thus perhaps easier to remember, four is four. Perhaps for this reason, children performed equally well when attributing statements using *two* and *four*.

Overall we conclude that the knowledge-state paradigm does not help children in drawing pragmatic inference with scalar terms, although numerals appear require separate consideration. In the following study we extend the paradigm to scales that are not quantifiers.

4. Study 3: Speaker Knowledge and Part/Whole Transformations

Recall that a scalar implicature is a pragmatic inference that compares what was said to a set of scalar alternatives. The current study asks participants to use alternative words that don't fall on a traditional scale, but would still follow Grice's expectation of describing something as fully as you know (quantity maxim). In the present experiment, two animal puppets watch the human character draw something. One animal then falls asleep while the other one remains awake. The human character then adds onto the original drawing, transforming it into something that belongs to a different category (e.g. transforming a cloud shape into ice cream on top of a cone; see Figure 4 for all transformation events). We expect that children will attribute statements showing full knowledge (i.e. including the new object) to fully informed speakers and statements showing partial knowledge (i.e. including original object) to partially informed speakers. The area of interest in this experiment is seeing whether children make the true-but-infelicitous choice of attributing statements with the original object (e.g. the cloud) to the fully-informed speaker who knows of the new object (i.e. the ice cream on a cone).

4.1 Participants

Participants were the total set of children and adults who took park in Studies 2 and 3. Overall this includes 28 children between the ages of 4;4 and 5;11 (mean age=5;0) and 16 adults between the ages of 18;8 and 21;8 (mean age = 19;7). The order of the studies was counterbalanced. No order effects were found for results of studies 2 and 3 so we will not discuss order further.

4.2 Materials and Methods

As with study 2, participants were removed from their classrooms with their assent and brought to a quiet room with a laptop computer. Children were again shown the two warmup videos and were corrected if they missed one. Only 3 participants (of 28) needed to be corrected after the first warm-up. All others passed both warm-up trials. After the 2 warmup videos, students were shown a series of 4 videos, as with Studies 2. Before each video, there would be a 10 second freeze frame of the video where the experimenter would introduce Sarah and the two animal puppets. The videos then started, showing Sarah drawing on a piece of paper while the two animal puppets "watched." Sarah would draw something on the piece of paper and then make it look as though she were finished drawing (e.g. putting the marker cap back on, looking satisfied, etc…). One of the animal puppets would then look bored and "fall asleep" by turning away from the action and lying down. Sarah would then cover the sleeping animal with a blanket. While the animal was sleeping, Sarah would motion to the other animal that she was going to continue drawing something. While the one animal was watching, she would add onto her original drawing, transforming it into something completely different from what she originally drew and belonging to a different semantic category. She then covered up her work with a large piece of fabric and woke up the sleeping animal. A second human character would enter the scene and ask the puppets, "What did Sarah do today?" Depending on which experimental condition the video was in, the children would hear one of the animals (sound file on the computer) say either "*Sarah drew a [the original object]*" or "*Sarah drew a [the modified object]*". The experimenter then asked the child, "*Which animal said that? Point to the animal who said that.*" Correct responses would match statements of partial knowledge (the original object) to the animal that fell asleep and complete knowledge (the transformed object) to the animal that remained awake.

As with Study 2, adult subjects were run in groups of 4. Each group of four sat in a small presentation room facing forward towards a projection screen that showed the same stimuli as the children. The adults did not attribute the statement verbally or by pointing, as the children did, but rather by silently circling a picture of the correct animal on their own response sheet.

4.3 Results

A one-way ANOVA reveals no significant effect of Statement (Strong, Weak) on children's performance (number of correct responses): $F(1,27)=.11$, $p=.09$. Children's performance on weak statements was not different from chance (=1.03 out of 2), but their responses to strong statements were different from chance (=1.42): $F(1,27)=3.05$, $p=.005$. Unlike Study 2, there was no effect of statement order (Strong/Weak first).

Individual differences played a moderate role. Out of the 28 children total, 10 were passers and 9 were failers. Of the 10 passers in the current study, 5 were also passers in their Study 2 (Quantifier or Verb) assignment. Of the 9 failers in the current study, 2 also failed their Study 2 assignment and 4 were at chance.

Adults performed as expected: 90.6% correctly matched the partial knowledge statements and 100% correctly matched the complete knowledge statements.

[INSERT FIGURE 6 HERE]

4.4 Discussion

Studies 1 and 2 show that children focus on the logical, semantic lower bound of weak scalar terms (e.g. *some*, *start*) and tend to extend the words into situations where they are still logically correct but underinformative. What is interesting is that this study did not require children to compute a traditional scalar implicature, as studies 1 and 2 did, yet children in this study performed similarly by attributing weak statements to fully informed speakers (true but infelicitous). For example, children were satisfied attributing the statement "Sarah drew a cloud" to a speaker who additionally knew that the cloud shape was currently part of an ice cream cone. Only 33.3% of 5-year-olds correctly matched the partial knowledge statement with the partially informed speaker (66.7% attributed to partial knowledge statement to the fully informed speaker - an infelicitous choice).

[INSERT FIGURE 5 HERE]

Scalar implicatures are one area where studies have experimentally shown that children tend not to use language pragmatically. The current data show that their semantic/logical bias extends into other areas of their language use.

5. General Discussion

The three experiments discussed here show three different ways of exploring children's pragmatic development. Study 1 showed that children are not very successful at pragmatically inferring relative quantities from *some* and *start* (they did so only 50% and 34.4% of the time, respectively). The remaining two studies similarly show that children are not particularly able to pragmatically infer that speakers who use weaker terms use them because they are not in a position (in this case, because of limited information) to use stronger terms. Study 2 shows specifically that our knowledge-state-paradigm prompts children to infer a pragmatic upper bound only 50% of the time for quantifiers such as *some*. Study 2 also shows that scalars can not be considered a homogeneous group. Numerals, for example, don't invite the same types of pragmatic inferences, presumably because they are typically understood to define an exact quantity, not just a logical lower bound, and might develop by different mechanisms (Hurewitz et. al., 2006). Children in this study correctly attributed partial information with number terms (e.g. statements with *two*) 64.3% of the time to a partially informed speaker. Finally, Study 3 shows that classical scalar implicatures are not the only area of language where children tend to accept true-but-pragmatically-infelitous information. In this study, they correctly attributed partial information to partially informed speakers only 53.57% of the time in situations that involved transforming a drawing of an object into a drawing of a different object.

The results of these studies are to be taken in perspective. As Noveck (2001) also concedes, the demands of the task are not to be dismissed and might, in fact, mask more ubiquitous pragmatic performance at a young age. This fact is particularly relevant when the results from different methodologies and paradigms are compared across the literature. Recall that Papafragou and Tantalou (2004), for example, found that 77.5% of time children can infer a pragmatic upper bound for a weak-SI, such as *some* when they simply have to use a scalar term to judge completion or non-completion (e.g. by awarding or withholding a prize for completion of an action to an animal character).

Children's performance in our experiments might improve if each child were to hear both statements with contrasting SIs (e.g. *some/all*, *two/four*) for each video and then assign a statement to each character. Perhaps hearing the contrast would bootstrap a pragmatic separation of *some* and *all*, for example. That is to say, if *some* and *all* are both available options, perhaps children would be better at deciding which one best fits each speaker.

While children's exact abilities regarding pragmatic inference seem to depend on the type of task, there is little doubt that children perform worse relative to adults. One explanation for this, presented in Noveck (2001), is that children first learn the logical, semantic interpretation of a word and only later in development learn the pragmatic limitations on its meaning. Perhaps alternatively, children have a different criterion for what is sufficient for conversation. Recall

Grice's quantity maxim (one should include as much information as is required). Perhaps children simply don't fully appreciate all that is required in a conversation and believe that a true-but-infelicitous statement is "good enough." This point is interesting when considering also that, developmentally, 5-year-old children have only recently (around age 4) acquired a theory of mind, a cognitive development that promotes children's abilities to reason metacognitively about the content of other minds. Perhaps a fledgling theory of mind does not fully equip children to reason about the demands of what another mind would need linguistically for precise communication. Perhaps this type of pragmatic inference relies on theory-of-mind abilities, which at age 5 are not fully developed. The connection between language and theory-of-mind is shown by Sabbagh and Baldwin (2001). They found that 4 year old children (more so than 3-year-old children) hesitate to learn a novel word from ignorant speakers, showing a sensitivity to speaker knowledge states. Sabbagh and Baldwin conclude that as children's theory-of-mind becomes more sophisticated, so too does their ability to reason about speakers. Clement, Koenig, and Harris (2004) similarly found that 4-year-olds (more so than 3-year-olds) monitor the accuracy and reliability of speakers in language acts.

One possible explanation for why children in the present studies performed as they did is that, while at age 5 they are beginning to develop the required metacognative abilities, they didn't have fully-developed capacity to succeed in the task, especially since our task was more demanding cognitively than the studies of Sabbagh & Baldwin (2001) and Clement, Koenig, & Harris (2004). Instead of having to implicitly connect a word to a referent on the basis of speaker reliability/knowledge, our participants had to explicitly judge which of two candidates is more likely to have spoken an utterance.

Horn (1972; as cited in Papafragou, 2006) delineates a stepwise rationale of what listeners compute implicitly when calculating a scalar implicature. Following Horn's example, the following shows the logic that one would need to implicitly access in order to succeed in our knowledge-state paradigm from Studies 2 and 3:

1. Someone uttered *"Sarah colored some of the stars"*

2*some* is lower-bounded by its semantic meaning (= 'at least one and possibly all')

3. There is a stronger statement, *all*, that could have been uttered which logically entails *some* but not vice versa.
4. There are two potential speakers present: Person A, who fell asleep partway through the event (partial knowledge), and Person B, who witnessed the full event (complete knowledge).
5. Given quantity requirements, Person B, who witnessed the full event, should report his full knowledge since it is relevant to the listener.
6. Person A, however, only has partial knowledge and knows for sure that only half of the event took place. Person A, assumed to be cooperative, should report his full knowledge (albeit incomplete) to the extent that he knows it to be true.
7. The statement *"Sarah colored some of the stars"* was NOT spoken by Person B since he would have communicated his full knowledge by using the stronger term *all*.
8. The speaker deliberately chose not to use the stronger term *all* because he was not in a position to do so.
9. Person A fits this criterion: he was not in a position to use a stronger term because he only witnessed the coloring of 2 starts out of 4. Person A uttered the statement.

In this way one (implicitly) computes a scalar implicature and places a pragmatic upper-bound on *some* to mean *not-all*. Children are not proficient at this implicature and though it is beyond the scope of our results to pinpoint which step is most problematic, it is reasonable to assume that children stumble at point 7; they do not seem as able as adults to compare what a speaker could have said to what they did say. It is logically correct to say *some* when one knows *all* to be true also. Children simply stop there and accept the *some* statement as good enough to fit either speaker.

We see from our adult subjects that by age 19 individuals become quite proficient at making pragmatic inferences from language. The question then becomes "how do pragmatic abilities develop over time?" Though it is beyond the scope of this paper to give a conclusive answer, it seems that a number of factors converge to produce implicature skills, including developing metacognitive abilities (e.g. theory of mind), as well as the additional experiences of learning the pragmatic nuances of word-choice from older, more sophisticated language users. After learning the logical, semantic domains of words, children begin to pick up on situations where the word is best used.

Acknowledgements

This research was supported by the Undergraduate Research Program at the University of Delaware. We would like to acknowledge Paul Quinn and Christine O'Hannessian for their helpful comments, and Sarah Lorch for her time and contributions. The research would not have been possible without the generous help from the parents, children, and administrators of Educare, Newark Day Nursery, TutorTime, the Lab School, and the Early Learning Center.

References

Clement F, M Koenig, and P Harris. (2004). The Ontogenesis of Trust. *Mind & Language* Vol. 19 No. 4, pp. 360–379.

Grice, H. P. 1989. Studies in the way of words. Cambridge, MA: Harvard University Press.

Horn, L. 1972. On the semantic properties of the logical operators in English. Doctoral diss., UCLA.

Hurewitz, F., A. Papafragou, L. Gleitman, and R. Gelman. Asymetries in the Acquisition of Numbers and Quantifiers. *Language Learning and Development* 2(2), 77-96

Noveck, I. 2001. When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition* 78, 165-188.

Papafragou, A. and J. Musolino. 2003. Scalar Implicatures: Experiments at the semantics-pragmatics interface. *Cognition* 86: 253-282.

Papafragou, A. and N. Tantalou. 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.

Pinker, S. 1994. The Language Instinct. New York: William Morrow.

Sabbagh, M. and D. Baldwin. (2001). Learning Words from Knowledgeable versus Ignorant Speakers: Links Between Preschoolers' Theory of Mind and Semantic Development. *Child Development* 72(4), pp. 1054-1070.

Shatz, M. (1980). Communication. In P. Mussen (Series Ed.) and J. Flavell, & E. Markman (Vol. Eds.), *Handbook of child psychology:* Vol 3. *Cognitive development* (pp.841-849). New York: Wiley.

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

Tables and Figures

Table 1: Materials for Study 1	
Quantifier Condition	**Verb Condition**
"Freddie Fish… "	*"Freddie Fish… "*
Test Items (Weak Scalars)	**Test Items (Weak Scalars)**
1)… "put **some** of the apples on the plate	1)… "**started** putting the apples on the plate
2)… "cut out **some** of the paper hearts	2)… "**started** planting the flowers"
3)… "planted **some** of the flowers"	3)… "**started** cutting out the paper hearts
4)… "put **some** of the coins in the bag"	4)… "**started** putting the coins in the bag"
Filler Items	**Filler Items**
5)… "put the bird in the birdhouse"	5)… "jumped on the table"
6)… "watered the plant"	6)… "watered the plant"
7)… "jumped on the table"	7)… "put the bird in the birdhouse"
8)… "turned on the flashlight"	8)… "turned on the flashlight"
Control Items (Strong Scalars)	**Control Items (Strong Scalars)**
9)… "opened **all** of the baskets"	9)… "**finished** opening the baskets"
10)… "turned on **all** of the lights"	10)…"**finished** turning on the lights"

TABLE 1: Materials for Study 1.

	Pass	Fail
Quantifier	4	4
Verb (note:2 other participants at chance	2	4

TABLE 2: Participants in Study 1 classified as strong "passers" or "failers." A classification of Pass or Fail is determined by 3 out of 4 responses as being passing or failing.

	1 action	2 actions	3 actions	4 (all) actions
Quantifier (*Some*)	12.5%	15.5%	25%	46.8%
Verb (*start*)	25%	9.4%	0%	65.6%

TABLE 3: Children's responses (between 1-4) on the Test Items (see Table 1) in Study 1.

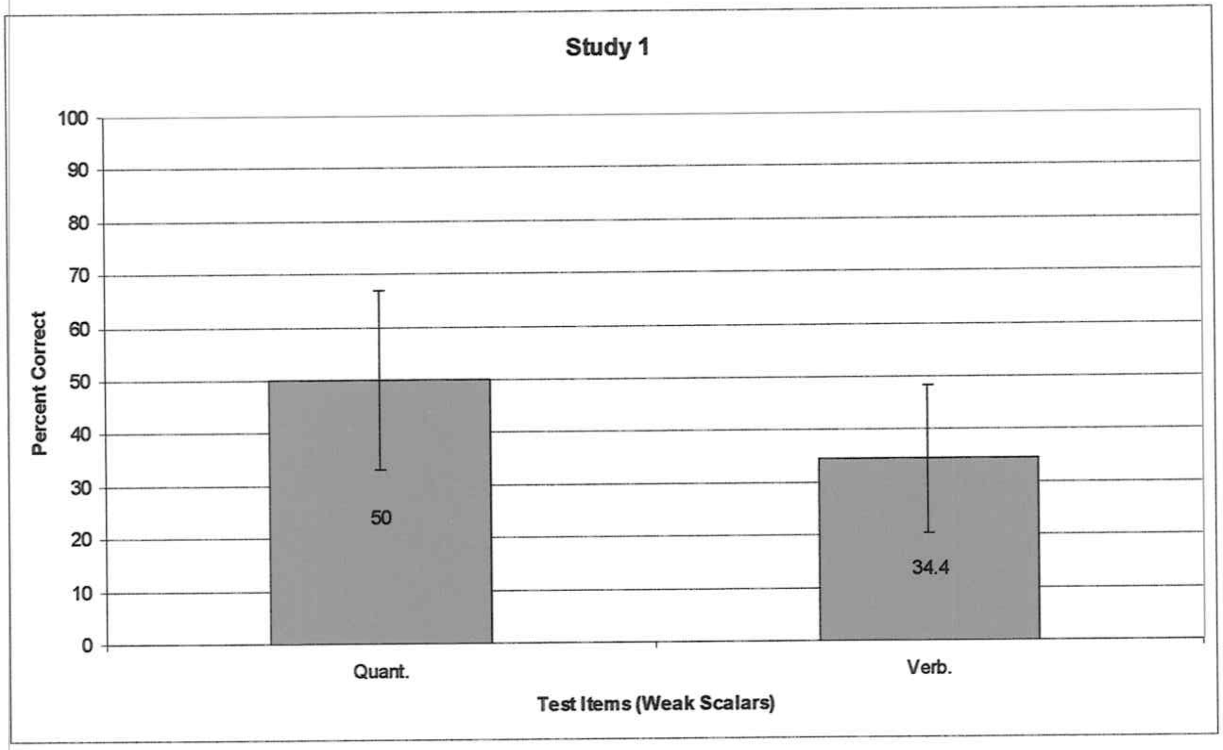


Figure 1: Children's responses to the Test Items (Weak Scalars) in Study 1: Quantifier Condition (*some*) and Verb Condition (*start*).

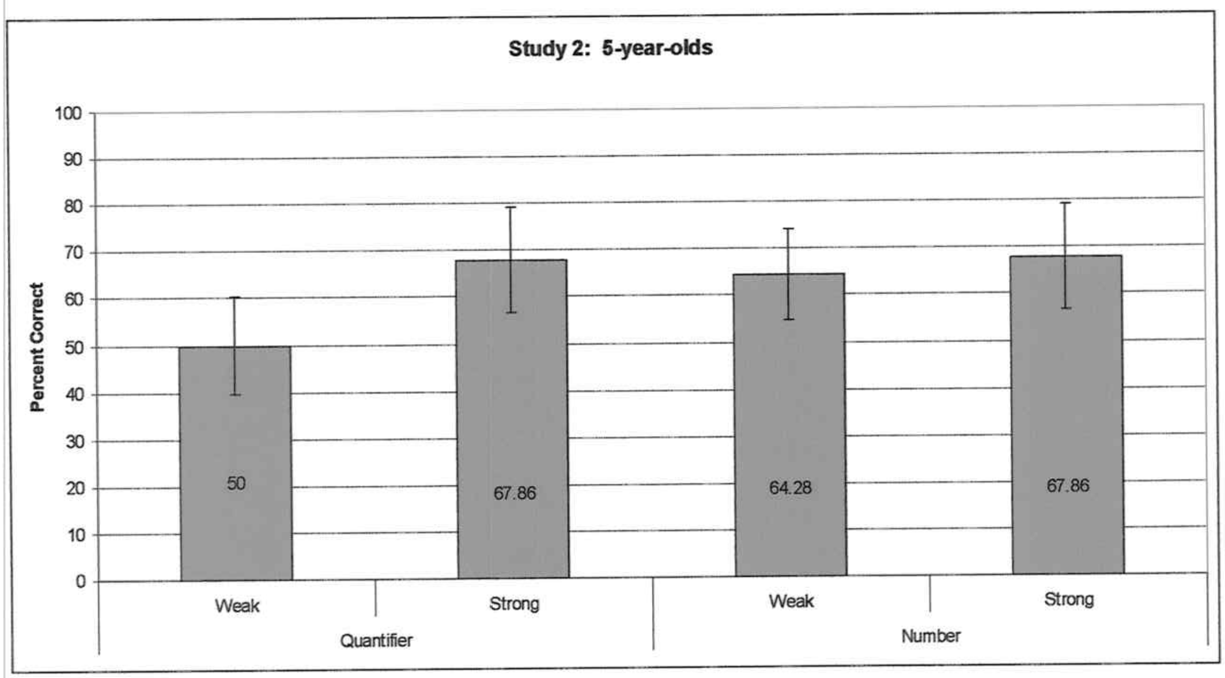


Figure 2: Children's performance in Study 2: Quant. Weak (*some*), Quant Strong (*all*); Number Weak (*two*) and Number Strong (*four*).

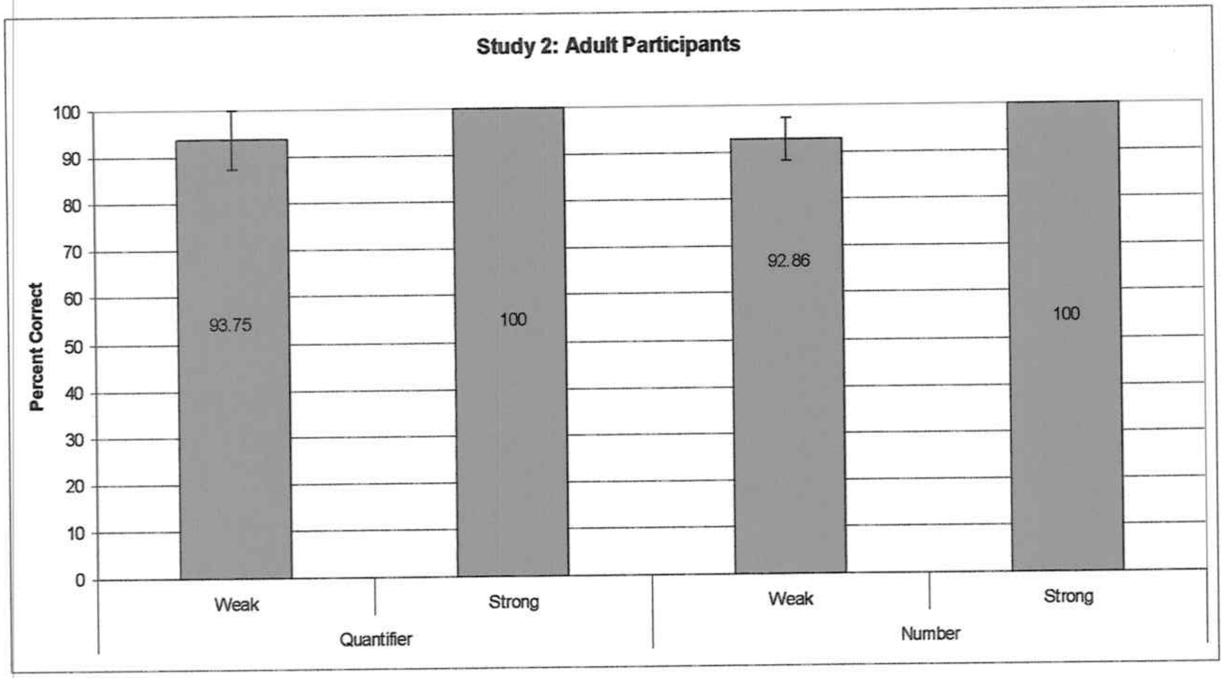
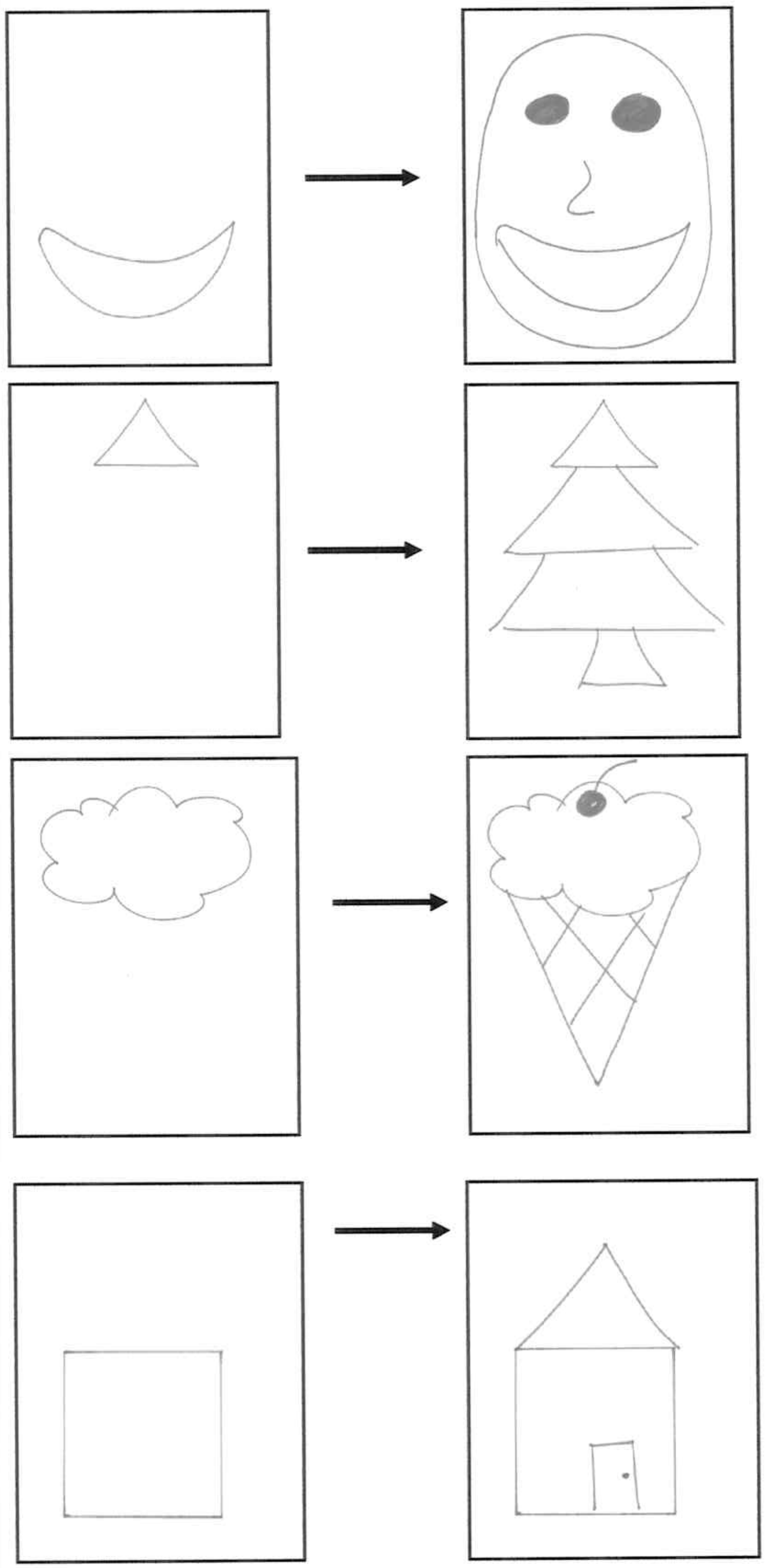


Figure 3: Adult Participants in Study 2: Quant. Weak (*some*), Quant. Strong (*all*); Number Weak (*two*), and Number Strong (*four*).

Figure 4: Transformation Events



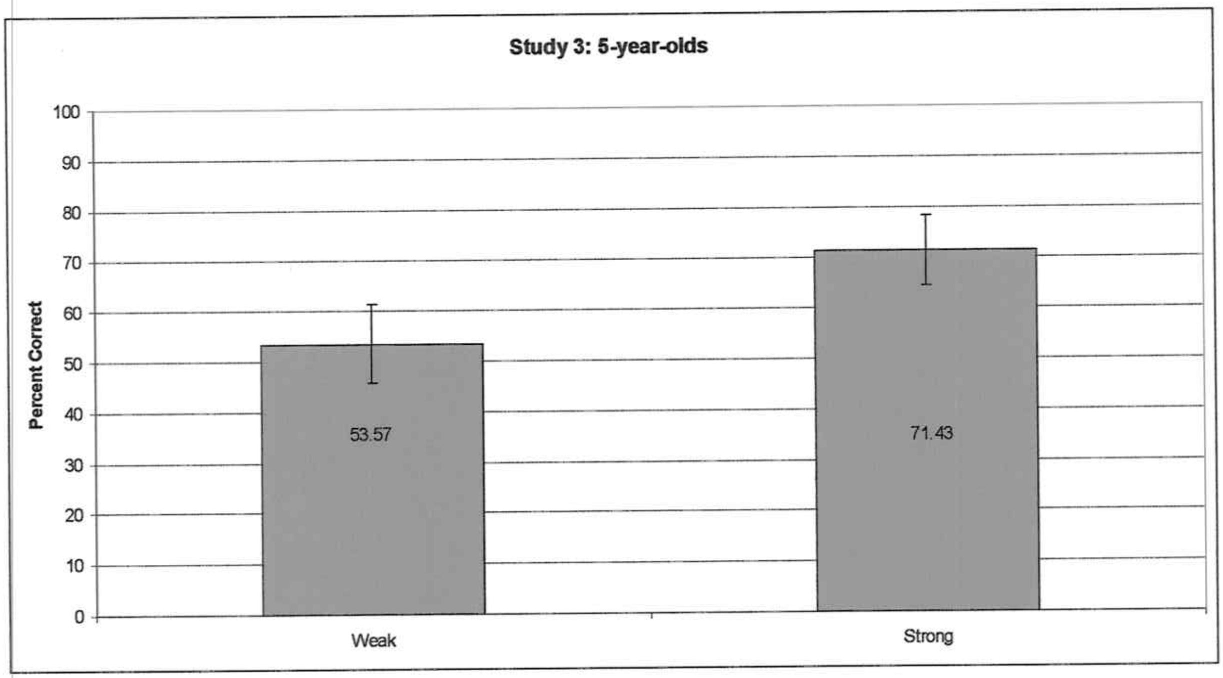


Figure 5: Children's performance in Study 3

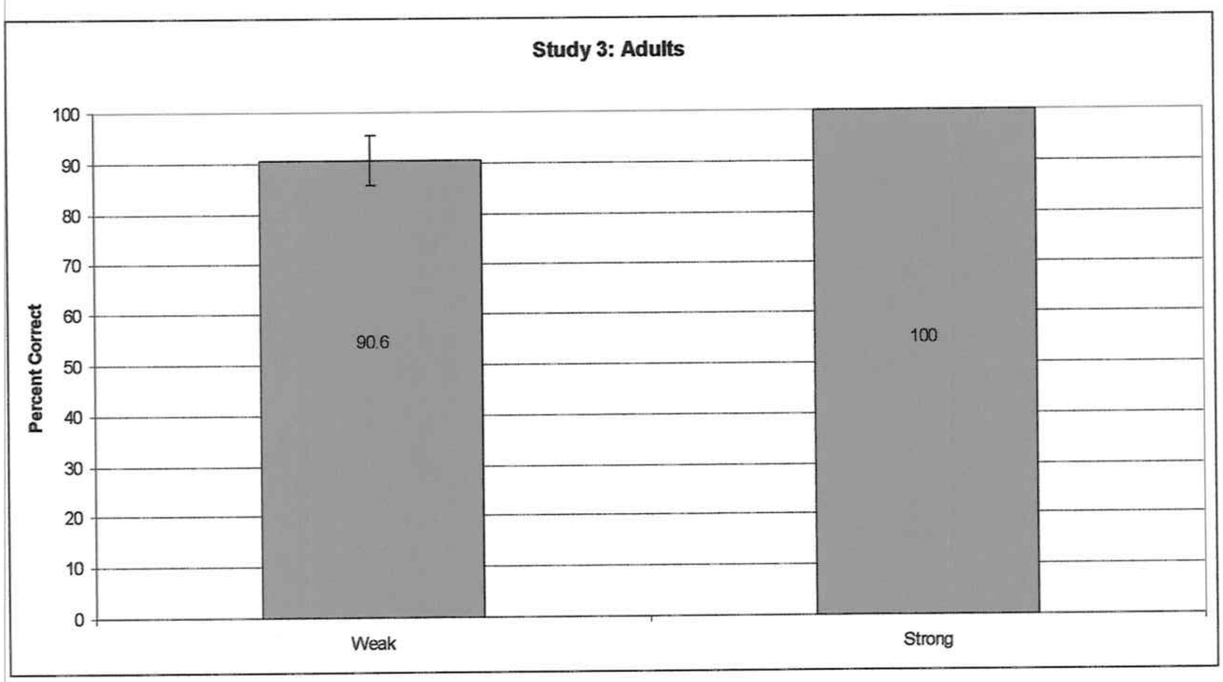
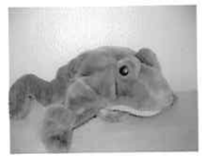


Figure 6: Adults' performance in Study 3.

Appendix: Sample Adult Response Sheet for Studies 2 and 3.

Participant Name ____________________
Participant Age ____________________
Date ____________________

1. Who said it? Circle one

2. Who said it? Circle one


3. Who said it? Circle one


4. Who said it? Circle one