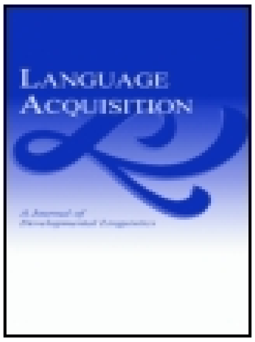

Language Acquisition

ISSN: 1048-9223 (Print) 1532-7817 (Online) Journal homepage: http://www.tandfonline.com/loi/hlac20

Children's Computation of Implicatures

Anna Papafragou & Niki Tantalou

LANGUAGE ACQUISITION, *12*(1), 71–82

SQUIBS AND REPLIES

Children's Computation of Implicatures

Anna Papafragou
Institute for Research in Cognitive Science
University of Pennsylvania

Niki Tantalou
Department of Cognitive Science
Johns Hopkins University

1. INTRODUCTION

This article focuses on a familiar kind of conversational implicature known as *scalar* implicature (SI):

(1) A: Do you like California wines?
B: I like some of them.
<u>Implicature: B doesn't like all California wines.</u>

Even though *some* is semantically compatible with *all*, it is used in (1) to communicate "some but not all." According to the traditional Gricean account of such examples, given that B could have used a more informative term (*all*) and as it would have been relevant to use *all* if it were true, A is entitled to infer that B is not, in fact, in a position to offer a statement containing *all*—most probably because such a statement is not true. Similar interpretations arise with logical connectives ("A or B" → *not A and B*), modals ("possibly x" → *not certainly x*), and a variety of other terms that can be seen to fall on an informational scale (see Grice (1989), Horn (1972; 1984); for varying perspectives, see Carston (1990; 1998), Chierchia (2001), Gazdar (1979), Levinson (2000), Sperber and Wilson (1986)).

Several studies show that preschool children have difficulties computing SIs. Chierchia, Crain, Guasti, Gualmini, and Meroni (2001) and Gualmini, Crain,

Requests for reprints should be sent to Anna Papafragou, Department of Psychology, 109 Wolf Hall, University of Delaware, Newark, DE 19716. E-mail: papafragou@psych.udel.edu

Meroni, Chierchia, and Guasti (2001) discovered that, in scenarios that made a statement containing the stronger term on a scale true (e.g., "Every boy chose a skateboard *and* a bike"), 5-year-old children—unlike adults—failed to reject a statement containing a weaker scalar term (e.g., "Every boy chose a skateboard *or* a bike"). In a related set of studies, Noveck (2001) found that 7- to 9-year-old children are more likely than adults to accept statements such as "Some giraffes have long necks," again presumably because they fail to generate the implicature <u>Not all giraffes have long necks</u> (cf. also Braine and Rumain (1981), Smith (1980)).

More recent work by Papafragou and Musolino (2003) showed that children's performance with implicatures improves under certain conditions. Their experiments included a training phase during which children were introduced to a puppet that occasionally said things that were "silly" (i.e., true but infelicitous). In the main phase, the children were asked to judge descriptions of acted-out stories given by the puppet: In one critical trial, a bunny completed a puzzle as part of a contest, but the puppet said that "the bunny did some of the puzzle." The introduction of training and of clear informativeness expectations in the critical trials made the children more likely to judge underinformative statements as bad descriptions of what happened. Nevertheless, even in these contexts, the children generally showed sensitivity to SIs only about half the time[1] (cf. also Papafragou (2003a)).

So far, the conclusion that children have limited sensitivity to scalar inferences has mainly relied on studies of logical expressions (e.g., quantifiers, connectives, modals) ordered in terms of entailment. It has long been known, however, that scalar inferences can be induced by any kind of salient contextual ordering, including stable (encyclopedic) or arbitrary (ad hoc) partial orderings (cf. (2) and (3), respectively) (Fauconnier (1975), Hirschberg (1985)):

(2) A: Have you read *A Beautiful Mind*?
B: I've read chapter 1.
<u>Implicature: B hasn't read the whole book</u>.

(3) A: Did you get an autograph from the Jacksons?
B: I got one from Janet.
<u>Implicature: B didn't get an autograph from Michael</u>.

A psycholinguistic account of scalar implicature needs to extend to cases such as (2) and (3) alongside the more familiar cases of entailment scales. At present,

[1]Certain scalar expressions seem to give rise to higher success with SIs: Numerals are one such case (Papafragou and Musolino (2003)), the degree modifier *half* is another (Papafragou (2003b)). There are good theoretical reasons for treating such "exact" numerical and degree modifiers differently from other scalars, which may explain children's exceptional performance (see Carston (1998), Horn (1992), and Papafragou (2003b) for discussion).

however, nothing is known about the derivation of such inferences by young children.

Furthermore, studies that document children's limited awareness of SIs have primarily relied on judgments of the acceptability[2] of "weak" scalar expressions in contexts in which a stronger term is warranted. These tasks, however, are different from the actual circumstances in which SIs are computed during naturalistic conversations in several respects. First, experimental conditions do not make it clear whether (or why) SIs should be considered as part of what the speaker actually *intended* to communicate. In ordinary cases of intentional communication such as (1), the speaker intends the addressee to conclude that the speaker does not like all California wines (and intends the addressee to recover this intention on the basis of what is said; cf. Grice (1989)). But in experimental designs used so far, the computation of SIs was not similarly constrained by the speaker's intention. In some experiments (Papafragou and Musolino (2003)), the speaker uttered an underinformative statement (probably out of incompetence) and may not have noticed it carries the potential for conveying an SI (cf. the "silly" puppet); in others (Noveck (2001)), underinformative statements were presented out of context and therefore invited participants to reconstruct a possible situation in which they could be uttered by an actual communicator. In short, previous tasks measured children's sensitivity to *potential* implicatures in an effort to approximate their performance with *actual* (communicated) implicatures.

Second, experimental scenarios in previous tasks did not raise specific-enough expectations of informativeness so as to motivate the computation of (even a potential) SI. In an exchange such as (1), A's question sets up relatively clear expectations of cognitive gains (i.e., to obtain information about B's opinion of California wines in general), which B's response then fails to meet. But such expectations were not consistently provided in test situations (e.g., the puppet was simply asked to report "what happened"; Gualmini et al. (2001), Papafragou and Musolino (2003)). In fact, when expectations of cognitive gains were made clearer and hence a stronger alternative became more salient, children's performance improved (Papafragou and Musolino (2003)).

Finally, previous tasks typically involved situations in which an utterance containing a scalar term (e.g., "Some of the Xs Ved") semantically conveys a true proposition (e.g., Some and possibly all of the Xs Ved) but carries a (potential) implicature that is false (Not all of the Xs Ved). To perform correctly in these tasks (i.e., to reject the statement), hearers should take the implicature (rather than simply the proposition expressed) as the basis for their assent/dissent with the

[2]Gualmini et al. (2001) and Chierchia et al. (2001) take scalar inferences to contribute to truth conditions (following Chierchia (2001)). They therefore interpret their participants' rejections of the weaker statements as truth-value judgments. We cannot go into the details of this theoretical approach here (see Horn (2003) for a defense of the Gricean, non-truth-conditional treatment of SIs). But we take our discussion of judgment tasks to apply also to truth-value judgment tasks, as they are partly designed in this case to detect effects of pragmatic inferencing.

original statement. To do so, participants had to estimate the experimenter's goals in setting up the task. This can be subtle: Adults, who are otherwise able to compute SIs, when presented with underinformative statements (e.g., "Some airplanes have wings") without supporting linguistic or extralinguistic context agree with the statements about half the time (Noveck (2001)).[3] More important, such judgments are removed from naturalistic conversations in which what is said and what is implicated are not normally pitted against each other but are taken jointly to contribute to what is meant by the speaker. Putting together the observations just presented, we conclude that the family of judgment tasks, however useful as an initial tool in exploring awareness of SIs, may in fact underestimate preschoolers' ability to compute implicatures "in the wild."

In this squib, we provide new experimental evidence on early implicature comprehension. Our investigation focuses on preschoolers' performance with SIs that are licensed by the familiar quantificational scale <*all, some*> as well as by encyclopedic and ad hoc partial orderings. Unlike previous studies, which have relied on truth value/pragmatic judgments, we introduce a new method that directly tests for the computation of SIs. This method reproduces, to the extent possible, the textbook environments for SIs given in (1) through (3). Our study has two main goals: first, to collect data about certain kinds of SIs that have so far escaped experimental attention and, second, to explore possible reasons for failures in early implicature-calculation in children. Specifically, if earlier reports of children's failures with SIs are (at least in part) due to the specific character of judgment tasks, there is an open possibility that children will be successful with scalar inferences from quantificational, encyclopedic, and ad hoc scalar orderings if a different method is used; in that case, there might be no difference in children's performance across the three types of scalar inference.

2. THE EXPERIMENT

2.1. Participants

Thirty Greek-speaking children participated in the study. They ranged in age from 4;1 to 6;1 (*M* age = 5;3). All children were recruited in a day care center in Athens.

2.2. Materials and Procedure

The children were presented with a set of animals and told that they would play a game. Each animal would be assigned a certain job. If at the end of the game the

[3]The training task in Papafragou and Musolino (2003) was designed to overcome this difficulty by showing that true but infelicitous statements should be rejected.

animal performed the job, the children should give the animal a prize; if not, the animal should get nothing. The children were randomly assigned to one of three conditions that corresponded to a certain scale type: quantificational, encyclopedic, or ad hoc.

The quantificational condition included the familiar <*all, some*> scale (Greek, <*oli, meriki*>). In a typical trial, an elephant was given a set of four paper stars and was told by the experimenter that he had to color them. The elephant then went into a dollhouse to do the coloring in quiet. After a while, he came out of the house, and the following conversation took place:

(4) Experimenter: Did you color the stars?
Elephant: I colored some.[4]

After hearing the animal's response the children had to decide whether the animal should receive a prize and to justify their response. We hypothesized that, if the children were able to compute SIs, they should refuse to award a prize to the elephant; furthermore, their justifications should reflect their sensitivity to the presence of the implicature. If children ignored SIs (e.g., if, in this example, they interpreted *some* as being compatible with *all*), they should be more generous in awarding prizes.

The encyclopedic condition included a set of orderings that were licensed by world knowledge (and supported by the visual context). For example, a bear had to eat a sandwich that consisted of bread, cheese, and ham. She decided to go into a nearby dollhouse so as not to litter the place with crumbs. When she came out of the house, the experimenter asked the following:

(5) Experimenter: Did you eat the sandwich?
Bear: I ate the cheese.

Finally, the ad hoc condition introduced a range of circumstantial, context-specific orderings. In one of the scenarios, a cow was assigned the task of wrapping two gifts: a toy parrot and a doll. Because she was too embarrassed to wrap them up in front of everyone, she decided to go into the dollhouse. When the cow reappeared, the experimenter asked the following:

(6) Experimenter: Did you wrap the gifts?
Cow: I wrapped the parrot.

This task offers a fairly straightforward means of evaluating children's pragmatic sophistication by making a certain behavior (here, the refusal to give a reward) contingent on the spontaneous computation of an implicature. The experimental

[4]Examples are translated from Greek throughout. In Greek, as in English, all conditions included contrastive stress on the scalar expression in the animal's response.

scenarios resemble naturalistic communicative circumstances in which implicatures are actually computed: SIs are part of what the animals intended to convey, and their recovery is set off by the fact that the animals failed to observe the required level of informativeness in answering the experimenter's question. Notice that the choice of the weaker alternatives in (4) through (6) has a natural motivation: The animals that were unable or unwilling to complete their task chose to report their partial progress (and only to imply that the task was not completed) in the hope of getting at least some reward.[5] For these reasons, the method presented here is an improvement on pragmatic judgment tasks previously used as a means of assessing early implicature calculation.

Children also received a number of control trials (which were identical across conditions and did not involve scalar expressions). In the control items, the animal characters always performed the action they had been assigned. Control items ensured that children could give positive (alongside negative) responses when asked whether an animal should be rewarded. A full list of experimental items is given in the Appendix.[6]

Ten children participated in each condition. In the quantificational group, children ranged from age 4;1 to 6;1 (*M* age = 5;3); in the encyclopedic group, children's ages ranged from 4;0 to 6;0 (*M* age = 5;4); in the ad hoc group, children ages ranged from 4;11 to 6;0 (*M* age = 5;2). In each condition, children received four control trials and four test trials administered in a pseudorandom order. Within each condition, order of presentation was counterbalanced among participants.

2.3. Results and Discussion

The overall result from the test trials is that children overwhelmingly refused to give a prize to the animal. Specifically, the children correctly withheld the prize 77.5% of the time in the quantificational cases, 70% in the encyclopedic cases, and 90% in the ad hoc cases. There was no reliable difference among these means, $F(2, 27) = 0.72$, $p = .49$. Overall, the children's performance in test trials was reliably different from chance responses ($p < .0001$).

[5]In the words of Larry Horn (2003), "it's safer to implicate the bad news rather than to assert it as part of what is said."

[6]One could argue that the children in the test trials might choose not to award a prize simply because the animals did not respond with a simple *yes*. We have two reasons to doubt that this could be true. First, we took care to ensure that animals *never* answered the experimenter's *yes/no* question with a simple *yes* or *no* but always offered a complicated/long response. In test trials, that response was equivalent to a *no*, whereas in control trials it was equivalent to a *yes*. Second, we have pilot evidence from similar experiments in English that shows that children can successfully reward the animals if completion of the desired action is part of an entailment ("Did you color some of the stars?" "I colored all of them"). Both these pieces of evidence show that children do not treat *all* indirect responses to *yes/no* questions as being equivalent to a *no*.

TABLE 1
Proportion of Correct Responses

Condition	*Test Trials*	*Control Trials*
Quantifier	77.5	97.5
Encyclopedic	70.0	100.0
Ad hoc	90.0	92.5

In control trials, the children were successful 97.5% in the quantificational condition, 100% in the encyclopedic condition, and 92.5% in the ad hoc condition. Again, no significant difference was found among these means, $F(2, 27) = 2.1$, $p = .14$. The children's performance in control trials was significantly different from chance ($p < .0001$). Results from the test and control trials are summarized in Table 1.[7]

After providing their responses, the children were asked to justify their answers. In cases where the children decided that a prize should be awarded, they justified this answer by stating that the animal had completed the action (or "had done its job," "had done what we told it to," etc.). More interesting, in cases where the children refused to award a prize, they always did so for the right reason, namely because they had inferred from the use of a "weaker" scalar term that a stronger term did not apply. Specifically, in the quantificational case, children in the majority (72%) of cases justified their negative responses by invoking the strong quantifier *all*: to use an example mentioned in the previous section, children refused to give a prize to an elephant that colored some of the stars because it did not color all of them. In the encyclopedic case, the children's justifications were more varied, with the most popular one involving the use of *only* (which surfaced 43% of the time). Finally, in the ad hoc condition, variability in the children's justifications increased (because the scalar ordering in these cases was neither as stable nor as transparent as in the previous two conditions), but their form again showed evidence for the successful retrieval of a scalar implicature. A full list of the types of justification offered by children on test trials is given in Table 2.

3. GENERAL DISCUSSION

These results provide compelling evidence for children's early ability to compute implicated aspects of a speaker's meaning. Specifically, we show that, in contexts

[7]Statistical comparison showed that the children were better on control than on test trials ($p = .02$). A closer look at the incorrect responses in the test trials reveals that certain children consistently failed the task, whereas others consistently passed. Specifically, "failers" include 4 children who failed all four test items and 2 children who failed three of four; "passers" include 3 children who passed three of four test items and 21 children who passed all four. Children who respond incorrectly in test items may be genuinely unable to detect SIs; alternatively, they may be overly lenient, awarding prizes even for incomplete actions. Our data do not allow us to adjudicate between these two possibilities at this point.

TABLE 2
Children's Justifications for Negative Responses on Test Trials

	Conditions		
	Quantificational	*Encyclopedic*	*Ad Hoc*
"He didn't do all the ——."[a]	72.0	25.0	11.0
"He only did ——."	—	43.0	17.0
"He didn't do the rest (too)."	—	7.0	19.0
"He didn't do —— [other part] (too)."	—	3.5	25.0
"He didn't do it (right)."	3.0	11.0	14.0
"He did some/a few."	12.5	3.5	—
Other	12.5	7.0	12.5

[a]Some justifications included a full verb rather than *do* ("He didn't clean all of ——," (etc.). We present all justifications schematically here for ease of exposition.

that approximate naturalistic conversations, children are capable of assessing informativeness expectations built during a talk exchange and of deriving SIs when these expectations are not met by the speaker's conversational contribution. The ability to assess expected levels of informativeness involves making spontaneous comparisons between a given linguistic stimulus (e.g., *some*) and other possible stimuli the speaker could have used to produce the cognitive effects she or he intended (e.g., *all*). This comparative ability is unambiguously demonstrated in participants' justifications for their negative responses, which regularly make reference to stronger alternatives ("He shouldn't get a prize because he didn't do ALL of X"). The massive success of the children in our study contrasts with results from previous studies, which have reported children's difficulties with SIs on the basis of truth-value/appropriateness judgment tasks and, in that sense, confirms the importance of task demands for children's performance with SIs (cf. Papafragou and Musolino (2003)).[8]

The structure of critical trials in our experiment included salient expectations of cognitive gains (set up by the experimenter's question) and highly accessible stronger alternatives to the weak scalars used by the animals. It would be worth investigating children's pragmatic sensitivity in circumstances where the cost of computing SIs becomes higher (e.g., where alternatives become less accessible by changing the structure of the experimenter's question or are altogether inferred from context). At present, the extent to which children can correctly attend to informativeness expectations—and monitor linguistic alternatives—in different contexts remains an open question. Nevertheless, other evidence suggests that the ability to consider contrastive alternates is not only within young learners' reach but is, in fact, very active in language acquisition. For instance, it is well known

[8]Based on our results, one may expect that children should stop short of computing SIs when informativeness expectations are satisfied by the lower scalar term. For some evidence in this direction, see Papafragou (2003a).

that observations of lexical contrast are instrumental at very early stages of word learning, where the fact that an adult used a novel word *a* rather than a (related) known word *b* in a certain context can be used by young learners to restrict the denotation of *a* to an appropriate *non-b* range (Carey (1978), Clark (1987)).

A striking aspect of our results is the finding that children's computations of informativeness are sensitive not only to logical entailment but also to stable ("encyclopedic") or arbitrary ("ad hoc") orderings. This is important because a large number of SIs in everyday conversation are of this more idiosyncratic variety. More generally, these findings raise the question of how children come to grasp a variety of other, nonscalar implicatures whose computation relies on idiosyncratic assumptions tied to specific contexts (cf. A: "Do you want to go to the movies?" B: "I'm tired" → <u>Implicature: B doesn't want to go to the movies</u>). The calculation of such inferences is a core task of the utterance interpretation device, and an account of how it is achieved is crucial for understanding how this device functions and grows.

Finally, evidence of children's comprehension of conversational implicatures may provide some insights into the architecture of the utterance interpretation system in adults. It has been proposed recently that certain types of scalar implicature belong to a class of default inferences that are supported by the presence of stable, contrastive alternates in the lexicon (Levinson (2000)). Such *generalized* conversational implicatures (GCIs) include the familiar quantificational, modal, and so on, scales and are assumed to arise by default, context independently, whenever an appropriate weak scalar element is present. These inferences are differentiated from *particularized* implicatures that depend on more circumstantial scalar orderings and hence rely heavily on context (cf. our encyclopedic and ad hoc scales). Even though this line of reasoning was not directly concerned with language development, one might expect that generalized SIs would prove easier for children to derive than particularized (context-specific) ones, other things being equal. For instance, it might seem plausible that the inference from *some* to *not all*—for the same reasons that it becomes generalized for adults—should be less computationally intensive for children compared to more ad hoc SIs. However, the results reported in this article do not offer evidence for such a split.[9] On the contrary, our findings can be used to support a unified view of implicature, according to which all implicatures are particularized and rely on context-specific computations of the speaker's communicative intentions.

ACKNOWLEDGMENT

Anna Papafragou is now with the Department of Psychology at the University of Delaware.

[9]For further arguments against GCI theory, see Carston (1990), Hirschberg (1985), and Geurts (1998).

REFERENCES

Braine, M. and B. Rumain (1981) "Children's Comprehension of 'or': Evidence for a Sequence of Competencies," *Journal of Experimental Child Psychology* 31, 46–70.

Carey, S. (1978) "The Child as Word Learner," in M. Halle, J. Bresnan, and G. Miller, eds., *Linguistic Theory and Psychological Reality*, MIT Press, Cambridge, Massachusetts.

Carston, R. (1990) "Quantity Maxims and Generalised Implicature," *UCL Working Papers in Linguistics* 2, 1–31.

Carston, R. (1998) "Informativeness, Relevance, and Scalar Implicature," in R. Carston and S. Uchida, eds., *Relevance Theory: Applications and Implications*, John Benjamins, Amsterdam, The Netherlands.

Chierchia, G. (2001) "Scalar Implicatures, Polarity Phenomena, and the Syntax/Pragmatics Interface," ms., University of Milan-Bicocca, Italy.

Chierchia, G., S. Crain, M. T. Guasti, A. Gualmini, and L. Meroni (2001) "The Acquisition of Disjunction: Evidence for a Grammatical View of Scalar Implicatures," in A. H.-J. Do et al., eds., *BUCLD 25 Proceedings*, Cascadilla Press, Somerville, Massachusetts.

Clark, E. (1987) "The Principle of Contrast: A Constraint on Acquisition," in B. MacWhinney, ed., *Mechanisms of Language Acquisition*, Lawrence Erlbaum Associates, Inc., Hillsdale, New Jersey.

Fauconnier, G. (1975) "Pragmatic Scales and Logical Structure," *Linguistic Inquiry* 6, 353–375.

Gazdar, G. (1979) *Pragmatics*, Academic Press, New York.

Geurts, B. (1998) "Scalars," in P. Ludewig and B. Geurts, eds., *Lexicalische Semantik aus kognitiver Sicht*, Narr Verlag, Tübingen, Germany.

Grice, P. (1989) *Studies in the Way of Words*, Harvard University Press, Cambridge, Massachusetts.

Gualmini, A., S. Crain, L. Meroni, G. Chierchia, and M. T. Guasti (200l) "At the Semantics/ Pragmatics Interface in Child Language," *Proceedings of SALT XI*, Cornell University, Ithaca, New York.

Hirschberg, J. (1985) *A theory of scalar implicature*, doctoral dissertation, University of Pennsylvania.

Horn, L. (1972) *On the Semantic Properties of the Logical Operators in English*, Doctoral dissertation, University of California, Los Angeles. [Distributed by IULC, Indiana University]

Horn, L. (1984) "Toward a New Taxonomy of Pragmatic Inference: Q- and R-Based Implicature," in D. Shiffrin, ed., *Meaning, Form, and Use in Context*, Georgetown University Press, Washington, D.C.

Horn, L. (1992) "The Said and the Unsaid," in C. Barker and D. Dowty, eds., *Proceedings of SALT II*, Linguistics Department, Ohio State University, Columbus.

Horn, L. (2003, June) "Grisotto Milanese," talk delivered at the Workshop on Scalar Implicature and Polarity, University of Milan-Bicocca, Italy.

Levinson, S. (2000) *Presumptive Meanings*, MIT Press, Cambridge, Massachusetts.

Noveck, I. (2001) "When Children Are More Logical Than Adults: Experimental Investigations of Scalar Implicature," *Cognition* 78, 165–188.

Papafragou, A. (2003a) "Aspectuality and Scalar Structure," *Proceedings from the 27th Annual Boston University Conference on Language Development*, Cascadilla Press, Somerville, Massachusetts.

Papafragou, A. (2003b) "Scalar Implicatures in Language Acquisition: Some Evidence From Modern Greek," *Proceedings from the 38th Annual Meeting of the Chicago Linguistics Society*, University of Chicago Press, Chicago, Illinois.

Papafragou, A. and J. Musolino (2003) "Scalar Implicatures: Experiments at the Semantics-Pragmatics Interface," *Cognition* 86, 253–282.

Smith, C. (1980) "Quantifiers and Question-Answering in Young Children," *Journal of Experimental Child Psychology* 30, 191–205.

Sperber, D. and D. Wilson (1986) *Relevance: Communication and Cognition*, Harvard University Press, Cambridge, Massachusetts.

APPENDIX

TABLE A1
Test Items (Translated from Greek)

Condition	*Animal's Task*	*Experimenter's Questions*	*Animal's Response*
Quantifier	Tiger has to eat four oranges	"Did you eat the oranges?"	"I ate some."
	Elephant has to color four stars	"Did you color the stars?"	"I colored some."
	Horse has to clean four toys	"Did you clean the toys?"	"I cleaned some."
	Pig has to feed four frogs	"Did you feed the frogs?"	"I fed some."
Encyclopedic	Bear has to eat a sandwich	"Did you eat the sandwich?"	"I ate the cheese."
	Frog has to paint a house	"Did you paint the house?"	"I painted the roof."
	Bunny has to clean a wagon	"Did you clean the wagon?"	"I cleaned the wheels."
	Elephant has to wash himself	"Did you wash yourself?"	"I washed my ears."
Ad hoc	Giraffe has to put four animals in a row (starting with cow)	"Did you put the animals in a row?"	"I put the cow." [lit.]
	Cow has to wrap two gifts (a toy parrot and a doll)	"Did you wrap the gifts?"	"I wrapped the parrot."
	Dog has to clean a merry-go-round (which has purple seats and one yellow seat)	"Did you clean the merry-go-round?"	"I cleaned the yellow seat."
	Lion has to feed snake (its food: a lollipop and a sausage)	"Did you give the snake its food?"	"I gave it the lollipop."

TABLE A2
Control Items (Translated from Greek)

Condition	*Animal's Task*	*Experimenter's Questions*	*Animal's Responses*
All conditions	Bear has to drink some milk	"Did you drink the milk?"	"I drank it and liked it."
	Worm has to read a book	"Did you read the book?"	"I read it. It was a nice story."
	Frog has to fix a broken chair	"Did you fix the chair?"	"I fixed it but it was hard."
	Starfish has to put flower in vase	"Did you put the flower in the vase?"	"In the beginning I couldn't find a vase but in the end I found it and put the flower in."