J. Child Lang. **33** (2006), 721–757.
doi:10.1017/S0305000906007550 Printed in the United Kingdom

From scalar semantics to implicature: children's interpretation of aspectuals*

ANNA PAPAFRAGOU
Department of Psychology, University of Delaware

(*Received* 10 *February* 2003. *Revised* 10 *August* 2005)

ABSTRACT

One of the tasks of language learning is the discovery of the intricate division of labour between the lexical-semantic content of an expression and the pragmatic inferences the expression can be used to convey. Here we investigate experimentally the development of the semantics–pragmatics interface, focusing on Greek-speaking five-year-olds' interpretation of aspectual expressions such as *arxizo* ('start') and degree modifiers such as *miso* ('half') and *mexri ti mesi* ('halfway'). Such expressions are known to give rise to scalar inferences cross-linguistically: for instance, *start*, even though compatible with expressions denoting completion (e.g. *finish*), is typically taken to implicate non-completion. Overall, our experiments reveal that children have limited success in deriving scalar implicatures from the use of aspectual verbs but they succeed with 'discrete' degree modifiers such as 'half'. Furthermore, children are better at spontaneously computing scalar implicatures than judging the pragmatic appropriateness of scalar statements. Finally, children can suspend scalar implicatures in environments where they are not supported. We discuss implications of these results for the scope and limitations of children's ability to both acquire the lexical semantics of aspectuals and to compute implicatures as part of what the speaker means.

[*] Portions of this material have been presented at the 27th Annual Boston University Conference on Language Development (2003), the 38th Annual Meeting of the Chicago Linguistics Society (2003), and the 26th Annual Penn Linguistics Colloquium (2003). I wish to thank Lila Gleitman, Chris Massey, Larry Horn, Laura Wagner, Ellen Prince, John Trueswell and Angeliek van Hout for comments and suggestions. Many thanks also to Niki Tantalou and especially Stathis Selimis for their help in conducting the experiments. Address for correspondence: Anna Papafragou, *Department of Psychology, University of Delaware*, 109 Wolf Hall, Newark DE 19716. fax: (302) 831 3645; tel: (302) 831 2752; e-mail: papafragou@psych.udel.edu

INTRODUCTION

One of the goals of theoretical linguistics is to account for the intuitions of native speakers about what sentences in their language mean. In order to accomplish this goal, most linguistic theories adopt a distinction between linguistically encoded ('semantic') and contextually inferred ('pragmatic') aspects of meaning. The assumption underlying this distinction is that linguistic meaning is abstract and underspecified in several respects but is routinely enriched or completed through powerful (albeit complex) inferential mechanisms. Those mechanisms in turn are taken to be constrained by expectations about communication (and general beliefs about the world) which reasonable speakers and hearers share with each other (Grice, 1989; Sperber & Wilson, 1986; Levinson, 2000). Drawing the semantics–pragmatics distinction in this way allows us to maintain an elegant and abstract semantics for natural language expressions, thereby simplifying the task of grammatical theory. At the same time, it offers a way of connecting abstract semantic meaning with the way people actually use language in communication – for instance, by helping explain how the same linguistic expression can give rise to different interpretations in different contexts. It is now widely recognized that truly explanatory accounts of semantic knowledge should be able to connect with models of the interpretive mechanisms responsible for language comprehension.

Within work on language acquisition, however, the study of children's semantic competence has mostly been carried out separately from the study of the development of performance systems. Some researchers, mostly within the generative framework, have intentionally chosen to focus on semantics in order to demonstrate that child language is constrained by innately specified grammatical knowledge. For this project, performance factors are not relevant, even though they need to be carefully controlled for (see Crain & Thornton, 1998). In other cases, the emphasis on word meaning has been a legitimate idealization of the complexities of the language-learning task. For instance, commentators agree that, in order to successfully acquire word meanings, ''the child must be equipped with a learning mechanism that constructs, tests and modifies semantic representations by comparing information about how [words] are used by speakers across speech events'' (Pinker, 1994, p. 379). But in studying the complexities of this learning mechanism, most approaches to word learning have abstracted away from the fact that the way 'words are used by speakers across speech events' may be systematically context-dependent or that words in use give rise to a host of conversational inferences.

A further reason for separating semantic from pragmatic development comes from the fact that estimates of early pragmatic sophistication have mostly been pessimistic: early work on a number of phenomena from speech acts to figurative language showed that very young communicators were egocentric, oblivious to other interlocutors' intentions, and hence insensitive to subtle pragmatic aspects of interpretation (for an overview, see Shatz, 1980). An implicit assumption which emerges from this literature is that pragmatic ('contextually enriched') interpretations are acquired at a later stage than 'pure' semantic meaning – hence it is possible to study the acquisition of semantics independently from children's pragmatic development.

Despite these prevailing assumptions, the study of the development and psychological status of pragmatic inference can be instructive for theories of word learning in a number of ways. First, it can simplify the task of word learning: given that words in context give rise to a host of conversational inferences, separating these inferences from linguistically encoded content leads to a simpler picture of what learners need to know about lexical semantics. Second, it can account for the systematic patterns of how words are used and interpreted in communication. Third, it may offer a unified picture of how the semantic and processing systems are organized in children and adults. Finally, information about how children navigate the semantics–pragmatics interface might function as a source of evidence for the semantics–pragmatics distinction in the adult grammar (e.g. by being a testing ground for competing semantic–pragmatic hypotheses). For all these reasons, the study of the development of the language interpretation device, including the computation of pragmatic inference, could usefully complement and inform the study of semantic development.

This approach raises a number of important questions about how the language interpretation system grows during development. Some of these questions concern the nature of early conversational inferences, their systematicity and robustness: do children compute pragmatic inferences? If not, how do they recover and become pragmatically savvy adults? Other questions concern the mechanisms whereby contextual inferences are computed by the language comprehension device: are the mechanics of language processing fundamentally the same or different across children and adults? And if one assumes continuity across the two populations, what aspects of parsing actually develop? These issues are far from theoretically or empirically settled; moreover, there was until recently lack of methodological tools specifically designed for the study of utterance interpretation in children (but see Trueswell, Sekerina, Hill & Logrip, 1999; Crain & Thornton, 1998).

This paper takes up the question whether children are sensitive to pragmatic aspects of interpretation. One of its goals is to throw some light onto the nature and properties of early implicatures, and hence to begin to answer the question of when, and most importantly how, children come to grasp the mechanics of human ostensive communication. A second, related goal is to examine how the computation of pragmatic inference builds on children's knowledge of lexical semantics. We focus on the class of aspectual expressions, whose semantic and pragmatic properties have been the topic of much discussion in the linguistic literature. We begin by laying out some background on aspectuals and motivating their choice as testing ground for the development of the semantics–pragmatics interface.

Aspectuality and scalar structure

Aspectuality refers to the linguistic encoding of the internal temporal composition of events. According to a semantic analysis which goes back to Horn (1972), various aspectual expressions such as completion/inchoative verbs (<*finish*, *start*>) and perfective/imperfective adverbials (<*completely*, *halfway*>) form SCALES, i.e. sets of alternates ordered on the basis of informational strength. Informational strength is defined in terms of the asymmetric entailment relationships which hold between structurally equivalent propositions containing members of the scale. For instance, *finish* is informationally stronger than *start* since (1) asymmetrically entails (2) (Horn, 1972):

(1) John finished eating the cake.
(2) John started to eat the cake.

On this account, weak scalar members such as *start* have a lower-bounded semantics (e.g. *start* means 'at least start and possibly finish'). In the right contexts, however, the use of such weak alternates typically generates the inference that a stronger (i.e. higher-ranked) member of the scale could not have been used. For instance, statements such as those in (3a–5a) are standardly taken to give rise to the inferences in (3b–5b):

(3) a. John started to eat the cake. →
    b. John didn't finish eating the cake.
(4) a. John ate the cake halfway. →
    b. John didn't eat the cake completely.
(5) a. John ate half of the cake. →
    b. John didn't eat all of the cake.

Such non-completion inferences from the use of aspectual/degree expressions are examples of SCALAR IMPLICATURES (SIs) and arise systematically and crosslinguistically.[1] Just like other kinds of conversational

[1] A similar effect is observed in the semantic opposition between imperfective and perfective aspect: Even though the imperfective is semantically compatible with both a bounded and an unbounded event (unlike the perfective which encodes boundedness), the use of the imperfective in certain contexts precludes the 'bounded' reading through a scalar inference (Smith, 1991). For instance, the imperfective utterance *John was eating the cake* can be used to implicate that John did not eat the cake. The further inferences

implicature, SIs can be removed in the presence of appropriate linguistic or extra-linguistic cues – i.e. they are CANCELLABLE:

(6) John started to eat the cake – in fact he finished within half an hour.

Furthermore, SIs are expected to survive across synonymous expressions of the same grammatical class and salience level – i.e. they are NON-DETACHABLE:

(7) a. John started/began to eat the cake. →
b. John didn't finish eating the cake.

According to the standard analysis, the computation of SIs is tied to Gricean considerations of informativeness or quantity, according to which participants in a conversation should make their contributions as informative as is required by the purpose of the talk exchange (Grice, 1989). In the case of (3a), the idea is that reference to a culminated event is more informative (and in most circumstances, will be more relevant) than reference to an event which is in some unspecified stage of completion. More specifically, upon hearing an utterance such as (3a), the hearer is entitled to reason as follows (Horn, 1972):

(8) i. *Start* is lower-bounded by its literal meaning (*start* = 'at least start and possibly finish').
ii. There is a stronger statement than (3a), namely (1), such that the latter unilaterally entails the former but not *vice versa*.
iii. Given quantity requirements, if the speaker knows that John finished eating the cake, and that it would be relevant to the hearer to know this fact, it would be misleading for her to tell the hearer that John started to eat the cake.
iv. The speaker is assumed to be co-operative and observe the conversational maxims (including quantity).
v. Therefore, the hearer infers that the reason the speaker chose not to express the stronger proposition in (3a) is that the speaker didn't know for a fact that it was true.
vi. The hearer infers that, for all the speaker knows, (1) is false, that is, John did not finish eating the cake (=(3b)).

On this neo-Gricean picture, even though a simple statement of the form *S*(*start*) is compatible with *S*(*finish*), it is usually interpreted in conversation so as to exclude it. Put differently, the lower-bounded meaning of *start to P* is typically upper-bounded by a conversational implicature ('start but not

from boundedness to completion and from unboundedness to non-completion are pragmatic and can be cancelled: *John ate the cake but didn't quite finish it* (cf. Hay, Kennedy & Levin, 1999).

finish P'). In this sense, *start* patterns alongside other scalars such as quantifiers or numbers, which also give rise to SIs (*Some/Two children left*→'Not all/no more than two children left'; for discussion, see Horn, 1972; Grice, 1989; Levinson, 2000; and for various alternatives to the neo-Gricean account, see Hirschberg, 1985; Carston, 1990; Sperber & Wilson, 1986; Kratzer, 2003; Chierchia, 2004).

Temporal/aspectual meaning is a particularly interesting domain for looking at how semantics and pragmatics work together to produce scalar interpretations in both children and adults. First, the acquisition of a verb such as *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 mixing has begun? The answer seems to depend on complex considerations, including intentionality, expectations about the normal course of events, etc. Furthermore, even though *John started to bake a cake* entails that John was engaged in an activity whose end point was to produce a cake, it does *not* entail that he baked a cake. But what makes an eventuality part of another eventuality in the extension of P in case there is no complete P-eventuality? This is a version of the famous IMPERFECTIVE PARADOX which has been discussed extensively in the semantic and philosophical literature (Dowty, 1979; Smith, 1991; Verkuyl, 1993).

Second, the scalar structure of aspectuality is internally rich: there is a strong intuition that the utterances in (3a–5a), just like those in (3b–5b), are inherently related to each other. At the core of this intuition lies the fact that, for TELIC predicates such as *eat the cake* (which denote temporally delimited events), the degree of completion of the event corresponds to the state of the affected object. If one has started eating the cake, it follows that one has eaten part of it; similarly, if one has eaten the cake halfway, half of the cake has been eaten. Such direct objects of the verb have been called INCREMENTAL THEMES (Dowty, 1991) and are taken to MEASURE OUT the event denoted by the main verb (Krifka, 1992; Verkuyl, 1993; Jackendoff, 1996).[2] The stable and systematic relation between the unfolding of an event and the change of state of the objects involved in the event permits the derivation of subtle inferences about event structure which relate members of the aspectual class across the nominal and verbal domains – even though the isomorphism is by no means perfect (*John ate the cake halfway* usually means that he ate half of it, while *John baked the cake halfway* does not

[2] These cases differ from ATELIC predicates such as *eat cakes* (which are temporally non-delimited), where the direct object which measures out the event is a nonbound entity itself.

necessarily mean that he baked half of it). From a learning perspective, understanding these correspondences is an important part of the task of grasping the nature of aspectuality.

Aspectuality and the development of the semantics–pragmatics interface

Earlier studies looking at the acquisition of aspect have mostly focused on grammatical aspect (e.g. the contrast between the perfective *John ate the cake* and the imperfective *John was eating the cake*: Wagner, 2002; van Hout, 2002; Kazanina & Phillips, 2003, and many others; cf. General Discussion) but have not been specifically concerned with scalar (pragmatic) effects. The study of most direct relevance to our present purposes was conducted by Papafragou & Musolino (2003) and looked at whether Greek-speaking five-year-olds computed scalar implicatures from a variety of expressions, including the aspectual verb *arxizo* 'start'. In that study, children were presented with a scenario in which a character completed an action (e.g. a Smurf completed a puzzle). A puppet then offered a description of the story outcome using an inchoative verb (*The Smurf started to do the puzzle*). Five-year-olds tended to accept this statement as a good description of the story ending. Adults, however, overwhelmingly rejected the target statement and offered a stronger scalar alternative as a more acceptable description of the story ending (*No, the Smurf* FINISHED *the puzzle*). Since the target statement is true but underinformative, children's behaviour was taken as evidence that children were not sensitive to 'non-completion' (scalar) inferences from the use of aspectual expressions.

These results confirm earlier findings which suggested that, even though adult communicators compute SIs regularly during utterance comprehension (Noveck, 2001; Noveck, Chierchia, Chevaux, Guelminger & Sylvestre, 2002), otherwise linguistically competent preschoolers are often insensitive to SIs. In one study, children and adults were presented with statements of the form *x might be y* in contexts in which the stronger statement *x must be y* were true (Noveck, 2001). It was found that 7 to 9-year-old children treated *x might be y* logically (i.e. as compatible with *x must be y*) much more often than adults. Similar results were obtained with quantifiers: 8 to 10-year-old children treated French *certains* ('some') as compatible with *tous* ('all') much more often than adults (Noveck, 2001; cf. Smith, 1980). In another study, children heard the statement *Every boy chose a skate-board or a bike* in a situation where each one of a set of four boys had chosen both a skate-board and a bike to play with. It was found that children failed to reject this description of the experimental scenario and to give their own corrected version of the story using *and* (Chierchia, Crain, Guasti, Gualmini & Meroni, 2001; Gualmini, Crain, Meroni, Chierchia & Guasti, 2001).

Children's behaviour in these tasks has been interpreted as an indication that children go through a stage during which they may not always attend to pragmatically derived aspects of meaning (Noveck, 2001) and/or compare more informative alternatives to the linguistic stimulus they are presented with (Chierchia *et al.*, 2001).

Several commentators have considered the possibility that children's failures in previous studies might be due to experimental demands rather than a genuine inability to compute scalar pragmatics. Preliminary support for this conclusion comes from a task in the Chierchia *et al.* (2001) study: if children are given a choice between a stronger and a weaker scalar (*Every farmer cleaned a horse or a rabbit* vs. *Every farmer cleaned a horse and a rabbit*) in a context in which the stronger one is most appropriate (e.g. where both a horse and a rabbit were cleaned), they correctly pick the most informative statement as a description of the story outcome. Papafragou & Musolino (2003) showed that, if task demands are clear and the informativeness expectations salient, 5-year-olds' ability to derive SIs improves. In a follow-up experiment, they trained children to correct a puppet who occasionally said 'silly' things (i.e. uttered true but infelicitous statements) and then asked children to judge descriptions of acted-out stories given by the puppet. In a modified version of one of the earlier stories, a bunny sets out to complete a puzzle and succeeds in doing so despite his friends' doubts. Asked specifically about the bunny's performance, the puppet offers an underinformative statement (*The bunny started to do the puzzle*). Under these conditions (training plus specific informativeness expectations built into the story), children are more likely to reject underinformative responses. Nevertheless, their derivation of implicatures is still poor: weak statements are rejected only about half of the time.

The acquisition of SIs raises a number of theoretical and methodological issues. One issue concerns children's ability to calculate implicatures: the literature converges on the conclusion that it is not adult-like but nevertheless leaves open the possibility that, in certain contexts and presentation conditions, children are successful at deriving SIs (see Noveck, 2001; Papafragou & Musolino, 2003; Papafragou & Tantalou, 2004 on the importance of task demands). Scalar inferences raise broader questions about children's language processing abilities. If children appear oblivious to SIs, what is responsible for such early failures? Is the architecture of children's processing system the same as that of the adult processor? Related questions are raised about the nature of early pragmatic inferences: are the scalar implicatures children draw cancelable? Are they non-detachable? A better understanding of the behaviour of these inferences is ultimately going to be important in describing the mechanisms that underlie their computation and use in discourse. Finally, one needs to understand more precisely the contribution of different scalar expressions to the derivation of SIs. Within semantic theory, there has been considerable discussion of the similarities and differences in the semantic contribution of different expressions which all seem to give rise to scalar effects (Sadock, 1984; Koenig, 1993; Carston, 1990; Horn, 1992; and *Asymmetries across scalars* below). It remains an open possibility that different scale types follow different acquisition paths and hence children come to grasp the scalar profile of different expressions in radically different ways.

The present study takes these earlier results as its starting point and seeks to investigate children's interpretation of aspectual expressions more systematically. The research reported here, conducted with speakers of Greek, includes a broader range of aspectual/degree expressions: the inchoative verbs *arxizo* ('start') and *ksekino* ('begin'), and the degree modifiers *miso* ('half') and *mexri ti mesi* (lit. 'until the midpoint', i.e. 'halfway'). In a series of experiments, we investigate the interpretation of these expressions in five-year-olds by looking at both their pragmatic properties (Experiments 1 and 2) and their lexical semantics (Experiment 3). The experiments also compare children's interpretations to the preferred pragmatic interpretations of these expressions by adults (Experiment 1). Our primary goal is to identify contexts and presentation conditions in which children might be successful at deriving SIs. Further, by focusing on a single semantic domain, we want to address the nature and properties of early pragmatic inferences (cancelability, non-detachability). Finally, by comparing children's sensitivity to the class of inferences in (3b–5b), we want to show how the study of scalar phenomena can throw light on the acquisition of aspect and reveal potential differences among the members of the aspectual class.

EXPERIMENT 1

METHOD

Participants

Participants in this study were a group of 40 Greek-speaking 5-year-olds between the ages of 4;10 and 5;11 (mean 5;6) and a group of 40 adult native speakers of Greek. Children participants were recruited from daycares in Athens, Greece. The adult speakers were also recruited from the Athens area.

Materials and procedure

The present study used a pragmatic judgment task to test whether children compute scalar inferences from the use of aspectual/degree expressions

TABLE 1. *Puppet statements on training stories in Experiment 1 (translated from Greek)*

Felicitous statements:	This is an elephant.	(target: an elephant)
	This is a frog.	(target: a frog)
Infelicitous statements:	This is something for eating food.	(target: a spoon)
	This is a little animal with four legs.	(target: a dog)

(for a discussion of the method, see Papafragou & Musolino, 2003). The main phase of the experiment was preceded by a training phase which aimed at making children familiar with the task of producing pragmatic judgments. Children were presented with a puppet, Minnie. They were told that Minnie would be shown some acted-out stories and then she would be asked what happened in the story. Children were told that Minnie sometimes says 'silly things' and that they should help her 'say things better'. In one of the training scenes, Minnie was shown a spoon and asked what it was. She described the object as 'something for eating food'. When asked whether Minnie answered well, children were expected to correct this truth-conditionally accurate but pragmatically infelicitous statement. Whenever they failed to do so, the experimenter eventually corrected Minnie and offered a more appropriate description of the object (*Minnie didn't say that very well. This is a* SPOON).

The training phase included two true but pragmatically inappropriate descriptions and two descriptions which were both true and appropriate. This was to make sure that children didn't develop a bias for assuming that Minnie always said silly things. The full set of training trials is given in Table 1.

In the main part of the experiment, children were shown a set of four test stories and four control stories. Each test story satisfied the semantic content of an informationally stronger element within an aspectual/degree scale but was described by Minnie in terms of a weaker element from that scale. For instance, in one of the test stories, Daisy watched Mickey while he carefully coloured a star. Daisy wanted to colour another star for herself, even though Mickey doubted that she could do it. At the end of the story, Daisy managed to colour the star. When asked how Daisy did, Minnie offered a statement such as the following:

(9) I Dezi arxise na zografizi to asteraki.
DET-Daisy started to colour DET-star-DIMIN.
'Daisy started to colour the star'.

(10) I Dezi ksekinise na zografizi to asteraki.
DET-Daisy began to colour DET-star-DIMIN.
'Daisy began to colour the star'.

https://doi.org/10.1017/S0305000906007550 Published online by Cambridge University Press

(11)	I Dezi	zografise	to	miso	asteraki.
	DET-Daisy	coloured	DET	half	star-DIMIN.
	'Daisy coloured half the star'.				
(12)	I Dezi	zografise	to	asteraki	mexri ti mesi.
	DET-Daisy	coloured	DET	star-DIMIN.	up-to DET-middle
	'Daisy coloured the star halfway'.				

After hearing Minnie's statement, children were asked whether Minnie had 'answered well'. In case they responded 'Yes', no further questions were asked. In case they responded 'No', children were asked whether we can 'say it better'. It was expected that, if children are pragmatically savvy, they should reject statements such as (9)–(12) as descriptions of the story and offer utterances such as (13)–(15) as improved ways of describing what happened:

(13)	I Dezi	teliose	to asteraki.	
	DET-Daisy	finished	DET star-DIMIN.	
	'Daisy finished the star'.			
(14)	I Dezi	zografise	olo to asteraki.	
	DET-Daisy	coloured	all DET star-DIMIN.	
	'Daisy coloured all of the star'.			
(15)	I Dezi	zografise	to asteraki	mexri to telos.[3]
	DET-Daisy	coloured	DET star-DIMIN	up-to DET-end
	'Daisy coloured the star to the end'.			

All test stories were identical across all conditions with the exception of Minnie's statements. All stories involved telic predicates (accomplishments) in which the endpoint of an event was specified through the direct object: specifically, we included two creation predicates, *ftiaxno to pazl* ('make the puzzle') and *xtizo ton pirgo* ('build the tower'); one transitive change-of-state predicate, *zografizo to asteraki* ('colour the star'); and one transitive activity verb with a locative modifier, *vazo tis kulures sto stilo* ('put hoops up around the pole'). Crucially, the progress of the event denoted by the predicate could be inferred by inspecting the state of the direct object. For instance, in order to determine whether Daisy had coloured the star (or had simply started colouring it), it was sufficient to inspect the proportion of the surface of the star which had colour on it. As a result, in all our stories, there was a mapping between the state of the object and the aspectual profile of the event as a whole, so that the direct object measured out the

[3] Greek verbs are obligatorily inflected for tense and aspect. There are two basic distinctions, past/nonpast (for tense) and perfective/imperfective (for aspect), which yield four possible tense-aspect combinations. The past-tense verbs in (13)–(15) are marked for perfective aspect.

TABLE 2. *Puppet statements in Experiment 1 (translated from Greek)*

Condition	Puppet statements on test trials
start	Mickey started to put the hoops up around the pole.
begin	Mickey began to put the hoops up around the pole.
half	Mickey put half the hoops up around the pole.
halfway	Mickey put the hoops halfway up around the pole.
start	The tiger started to build the tower.
begin	The tiger began to build the tower.
half	The tiger built half the tower.
halfway	The tiger built the tower halfway.
start	Daisy started to colour the star.
begin	Daisy began to colour the star.
half	Daisy coloured half the star.
halfway	Daisy coloured the star halfway.
start	The little girl started to make the puzzle.
begin	The little girl began to make the puzzle.
half	The little girl made half the puzzle.
halfway	The little girl made the puzzle halfway.
All Conditions	**Puppet statements on control trials**
	The little girl jumped over the fence.
	The giraffe ate the biscuit.
	The green frog caught the squid.
	The horse beat the tortoise (in a race).

event denoted by the main verb (see earlier discussion in Introduction).[4] The full set of test stories is given in Table 2.

Control items (which were identical across conditions) also involved two characters engaged in some sort of contest. They were always correctly (and felicitously) described by Minnie and never involved the use of scalar terms (see Table 2). For instance, in one of the fillers a horse and a turtle took part in a race and the horse finished first. Minnie's description was: 'The horse beat the turtle'. It was expected that children should find no difficulty accepting all control statements as good answers.

Children were randomly assigned to one of four conditions, which corresponded to the four scalar expressions used. Ten children participated in each condition. The age of the children in the 'start' condition ranged from 4;10 to 5;11, in the 'begin' condition from 4;11 to 5;11, in the 'half' condition from 5;0 to 5;9 and in the 'halfway' condition from 4;11 to 5;11; within each condition, the mean age of participants was 5;6. In

[4] In the case of the activity V+locative PP (e.g. 'put hoops around the pole'), measuring out is done through a Path which corresponds to the proportion of pole covered by the hoops. On the several types of measuring out, see Dowty (1991), Jackendoff (1996).

https://doi.org/10.1017/S0305000906007550 Published online by Cambridge University Press

each condition, subjects received four test trials and four control trials administered in a pseudo-random order. Within each condition, order of presentation was counterbalanced among subjects.

Adult participants were randomly assigned to one of the four conditions in a modified version of the same task. They were given a leaflet which contained in written form the instructions verbally given to the children. For the training, control and test trials, adults read a description of the stories which did not contain any scalar items and which did not specify the story's ending. For instance, for the star-colouring story, adults read the following description (translated from Greek):

(16) Daisy and Mickey are playing with markers. Mickey shows Daisy a star and demonstrates how to colour it. He tells her that one should not leave white spaces and makes it clear that he doesn't trust her to colour a star on her own. Daisy thinks she can manage. She takes another star and a marker. The story ends as shown in picture [number of picture provided].

In order to see how the story ended, adults had to consult a photograph in the booklet which showed the outcome of the event (in the above example, Daisy holding a coloured star). Participants then read Minnie's statement and had to answer the same questions as the children did ('Did Minnie answer well? If not, can we say it better?') by filling in their answers in the space provided. It was expected that adults would overwhelmingly correct the puppet's statements on all test trials.

In sum, we have a 2×4 experimental design where age and expression type are treated as between subjects factors.

Results

Our dependent measure was children's 'No' responses to the puppet's statements, i.e. their rejections of the scene descriptions offered by Minnie. The proportion of 'No' responses was entered into an ANOVA with two factors, age (5-year-olds, adults) and expression ('start', 'begin', 'half', 'halfway'). The analysis revealed a significant main effect of age ($F(1, 72) = 58.07$, $p < 0.0001$) but no significant main effect of expression ($F(3, 72) = 1.09$, $p = 0.35$) and no reliable interaction between age and expression ($F(3, 72) = 2.27$, $p = 0.08$).

Beginning with test trials, it was found that adult subjects overwhelmingly rejected the puppet's statements in all four conditions (92.5% of the time for 'start', 97.5% of the time for 'begin', 90% of the time for 'half' and 97.5% of the time for 'halfway'). Statistical analysis revealed no reliable difference among these rejection rates ($F(3, 36) = 0.61$, $p = 0.61$). By contrast, it was found that 5-year-olds did not generally reject the

https://doi.org/10.1017/S0305000906007550 Published online by Cambridge University Press

puppet's statements in the test trials (32·5% of answers were rejections for 'start', 27·5% for 'begin' and 35% for 'halfway'), with the exception of 'half' (67·5% of the answers were rejections). The difference among these means is not significant ($F(3, 36)=1{\cdot}8$, $p=0{\cdot}16$). Further analyses revealed that none of these means is significantly different from chance ('start': $p=0{\cdot}19$; 'begin': $p=0{\cdot}20$; 'half': $p=0{\cdot}24$; 'halfway': $p=0{\cdot}22$).[5]

On the control items, adults in the 'start' group gave correct answers 77·5% of the time, in the 'begin' group 82·5% of the time, in the 'half' group 97·2% of the time and in the 'halfway' group 87·5% of the time. There is no reliable difference among these means ($F(3, 36)=2{\cdot}05$, $p=0{\cdot}12$). Adults' 'wrong' answers on control items simply reflect different preferences for verbally describing the outcome of the story. For instance, in the story in which the horse beat the turtle in a race, adults who judged the statement 'The horse beat the turtle' as unsatisfactory proposed different variants (e.g. 'The horse reached the log first'). On the same items, children in the 'start' condition gave correct responses 97·5% of the time, in the 'begin' condition 90% of the time, in the 'half' condition 100% and in the 'halfway' condition 90% of the time. No statistical difference was found among these means ($F(3, 36)=0{\cdot}70$, $p=0{\cdot}55$).

Recall that, in case they rejected Minnie's statement, subjects had to offer an improved version of her utterance. It was expected that adults would not only reject Minnie's underinformative statements on the test trials but also go on to reformulate her descriptions using a stronger (more informative) term. This expectation was borne out: all corrections provided by adult subjects were of this sort.

Unlike adults, 5-year-old children generally accepted Minnie's statements for all scalar expressions (with the single exception of 'half'). Nonetheless, in those cases where children did reject the puppet's statement, they did so for the right reason, since their corrections invoked a more informative item than the one used by Minnie. Such adult-like corrections accounted for 91% of the children's corrections across all four conditions.

DISCUSSION

These findings confirm – and empirically extend – previous experimental reports that scalar inferences are regularly computed, when appropriate, during adult language comprehension but do not surface consistently in language processing by young children (Noveck, 2001; Chierchia *et al.*,

[5] We also split participants into Passers (children who gave correct responses on 3 or 4 out of the 4 test trials) and Failers (children who never or only once answered correctly in test items). For 'start' and 'halfway', we found 3 Passers and 6 Failers; for 'begin', 3 Passers and 7 Failers; but for 'half', we found 6 Passers and 3 Failers. (The remaining children gave 2 out of 4 correct responses.)

2001; Gualmini *et al.*, 2001; Papafragou & Musolino, 2003; among others).[6] Pragmatic failures occurred even in circumstances in which (a) children were trained to recognize and correct pragmatic infelicity; (b) the experimental demands made the stronger term on the scale highly relevant for the purposes of the exchange. An intriguing aspect of the present findings is the fact that, even though children generally have limited sensitivity to the pragmatics of scalar terms such as 'start' and 'halfway', there are differences between individual scalar terms (even though they don't reach significance levels). We return to this issue later on in the paper.

A possible concern with the present experimental design is that the derivation of SIs is removed from the contexts in which it most naturally occurs during actual conversations. Notice that, in order to pass this task, children need to perform a pragmatic judgment by explicitly comparing a certain linguistic stimulus (Minnie's description) to other potential stimuli which could have been produced in the given context. This is different from situations in which SIs are computed during naturalistic conversations in several respects. First, it is not clear that the specific SIs targeted in this experiment are actually part of what Minnie MEANT by her utterance. In ordinary cases of intentional communication, the speaker intends the addressee to compute implicatures (and he also intends the addressee to recover this intention on the basis of what is said; cf. Grice, 1989). But in our experiment, the computation of SIs was not similarly constrained by the speaker's intention: in fact, Minnie consistently uttered underinformative statements probably out of incompetence ('silliness') and may not have noticed that such statements carry the potential for conveying a SI. This task thus measured children's sensitivity to *potential* (albeit unwanted) implicatures in an effort to approximate their performance with *actual* (communicated) implicatures. But detecting potential implicatures does not involve the same machinery as computing intended ones (cf. Sperber & Wilson, 1986), and even though adults can do both, children could definitely be more successful in the latter task than in the former.

Second, experimental scenarios in this study may not have raised specific enough expectations of informativeness, so as to motivate the computation of (even a potential) SI. Actual conversations involve constant monitoring of expected cognitive advances (or gains, see Sperber & Wilson, 1986) and this involves the comparison of the actual linguistic stimulus to implicit alternatives which the speaker could have used but chose not to. The test situations did not consistently provide clear cognitive expectations, since the puppet was simply asked to report on the performance of the story protagonist. And even though our aim was to make the stronger scalar

[6] Papafragou and Musolino (2003) report higher success rates with 'start' using the same method but a slightly older population of 5-year-olds.

alternative (e.g. 'finish') highly relevant, the fact that such alternatives were not embedded in the conversational goals may have made them less obvious to young participants. Contrast this with a more realistic conversational exchange, where the speaker sets up relatively clear expectations of cognitive gains which may or may not subsequently be met by the addressee (e.g. Wife: *Did you make dinner?*, Husband: *I started making it*).

Finally, our task (as almost all previous tasks in the literature) involved situations in which an utterance containing a scalar term (e.g. *X started to V*) semantically conveyed a true proposition (e.g. 'X started and possibly finished Ving') but carried a (potential) implicature which is false ('X did not finish Ving'). In order to perform correctly in these tasks (i.e. reject the statement), hearers should take the implicature (rather than simply the proposition expressed) as the basis for their assent/dissent with the original statement. To do so, participants had to estimate the experimenter's goals in setting up the task: one source of evidence for reconstructing the experimenter's intent was the training task which showed that true but infelicitous statements should be rejected (cf. Papafragou & Musolino, 2003). However, despite our efforts to the contrary, children may still have chosen to formulate their responses on the basis of whether the puppet's statements were true or false (rather than felicitous or infelicitous). That is, children (unlike adults) may have ignored the moral of the training phase and may have gone on to accept statements which were simply true. In any case, these experimental circumstances are removed from naturalistic conversations, where what is said and what is implicated are not normally pitted against each other but are taken to jointly contribute to what is meant by the speaker.

Putting together the above observations, we conclude that this task, and the family of pragmatic judgment tasks more generally, however useful as an initial tool in exploring awareness of SIs, may not in fact provide an accurate picture of young children's ability to compute implicatures in everyday conversational exchanges. This raises the question whether children might be more successful with SIs overall if the methodological requirements of the task were different. This possibility was pursued and explicitly tested in Experiment 2.

EXPERIMENT 2

METHOD

Participants

Participants in this study were a new group of 40 Greek-speaking 5-year-olds between the ages of 5;1 and 6;3 (mean 5;6). Children were recruited from daycares in the same Athens area as in Experiment 1.

Materials and Procedure

The experiment involved a task in which a set of animals decided to build a school for young animals. Their model was a school for (human) children which was found in the experimental scene. For the project to be completed, each animal was assigned a certain task (e.g. to build the school building, bring in books, find a school sign, etc.). Children were instructed to give the animal a prize if, at the end of the game, the animal had performed the assigned task; otherwise, the animal should get nothing. For instance, the horse was asked to build the school. After going away for some time, the animal was brought back. The experimenter then said she would find out what the animal had done and went on to ask the horse:

(17) Extises to sxolio?
built-2ndper DET school
'Did you build the school?'

The horse then replied with one of the following (note that scalar elements bear contrastive stress):

(18) Arxisa na to xtizo.
started-1stper to it-CLITIC build
'I started to build it.'

(19) Ksekinisa na to xtizo.
began-1stper to it-CLITIC build
'I began to build it.'

(20) Extisa to miso.
built-1stper DET half
'I built half of it'.

(21) To extisa mexri ti mesi.
it-CLITIC built-1stper up-to DET middle
'I built it halfway'.

After hearing the animal's response, children had to decide whether the animal should receive a prize or not and justify their response. It was hypothesized that, if children were able to compute SIs, they should refuse to award a prize to the horse; furthermore, their justifications should reflect their sensitivity to the presence of the implicature.

This task offers a fairly straightforward means of evaluating children's pragmatic sophistication by making a certain behaviour (here, the refusal to give a reward) contingent on the spontaneous computation of an implicature. This method differs from Experiment 1 in several respects. First, and crucially, the child is no longer required to judge whether an utterance is acceptable. The child is simply asked to listen to a

https://doi.org/10.1017/S0305000906007550 Published online by Cambridge University Press

conversational exchange and interpret utterances in the usual way. Second, the exchange introduces clear and salient communicative expectations (cf. the *Yes-No* question with a telic predicate). Finally, and most importantly, the speaker now clearly intends to *communicate* the scalar (non-completion) implicature. Specifically, the use of a weaker scalar term (and the presence of an implicature) has a natural motivation: animals which were unable or unwilling to complete their task chose to report their partial progress (and only imply that the task was not completed). That is, the use of a scalar expression is motivated by the clash between the demands of informativeness (which require that the question to the animal be answered appropriately) and truthfulness (which require that the animal does not say something which it knows to be false). In all these respects, the new experimental scenarios resemble naturalistic communicative circumstances in which implicatures are actually computed and are thus an improvement over pragmatic judgment tasks previously used as a means of assessing early implicature calculation (cf. Papafragou & Tantalou, 2004 for related discussion).

Children also received a number of control trials (which were identical across conditions and did not involve scalar expressions). In control items, the animal characters always reported that they performed the action they had been assigned. As with test items, animals avoided a *Yes/No* response but gave longer answers. 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 Table 3.

Children were randomly assigned to one of four conditions, which corresponded to the four scalar expressions used. As in Experiment 1, test stories involved telic predicates. Ten children participated in each condition. In the 'start' group, children ranged from 5;1 to 5;8 (age mean: 5;6); in the 'begin' group, children's age ranged from 5;3 to 5;8 (mean: 5;6); in the 'half' group, children ranged from 5;3 to 5;8 (mean: 5;6); finally, in the 'halfway' group, children ranged from 5;5 to 6;3 (mean: 5;8). In each condition, subjects received four test trials and four control trials administered in a pseudo-random order. Within each condition, order of presentation was counterbalanced among subjects.

RESULTS

As before, children's correct ('No') responses on test trials were entered into an ANOVA with expression ('start', 'begin', 'half' and 'halfway') as a between-subjects factor. The analysis revealed a main effect of expression ($F(3, 36) = 4{\cdot}48$, $p = 0{\cdot}0089$). Specifically, 5-year-olds correctly refused to

TABLE 3. *Animal statements in Experiment 2 (translated from Greek)*

Condition	Animal statements on test trials
	(Horse has to build the house)
	Experimenter to horse: Did you build the house?
start	– I started to build it.
begin	– I began to build it.
half	– I built half of it.
halfway	– I built it halfway.
	(Cat has to make the school sign)
	Experimenter to cat: Did you make the sign?
start	– I started to make it.
begin	– I began to make it.
half	– I made half of it.
halfway	– I made it halfway.
	(Dog has to weave a door mat)
	Experimenter to dog: Did you weave the mat?
start	– I started to weave it.
begin	– I began to weave it.
half	– I wove half of it.
halfway	– I wove it halfway.
	(Bear has to draw a star for the school decoration)
	Experimenter to bear: Did you draw the star?
start	– I started to draw it.
begin	– I began to draw it.
half	– I drew half of it.
halfway	– I drew it halfway.
All conditions	**Animal statements on control trials**
	(Pig has to carry books for the school library)
	Experimenter to pig: Did you bring the books?
	– I brought them, even though they were heavy.
	(Lion has to plant a tree)
	Experimenter to lion: Did you plant the tree?
	– I planted it and can show it to you.
	(Frog has to bring a chair for the teacher to sit)
	Experimenter to frog: Did you bring the chair?
	– I found a nice one and brought it.
	(Turtle has to put up a fence)
	Experimenter to turtle: Did you put up the fence?
	– It took me time but I put it up.

give a prize in the test trials 47·5% of the time for 'start', 47·5% of the time for 'begin', 75% of the time for 'halfway', and 100% of the time for 'half'. Pairwise comparisons between means reveal significant differences only for 'half' vs. 'start' ($p=0{\cdot}002$), 'half' vs. 'begin' ($p=0{\cdot}002$) and 'half' vs. 'halfway' ($p=0{\cdot}02$). Children's performance in the 'start' and 'begin'

https://doi.org/10.1017/S0305000906007550 Published online by Cambridge University Press

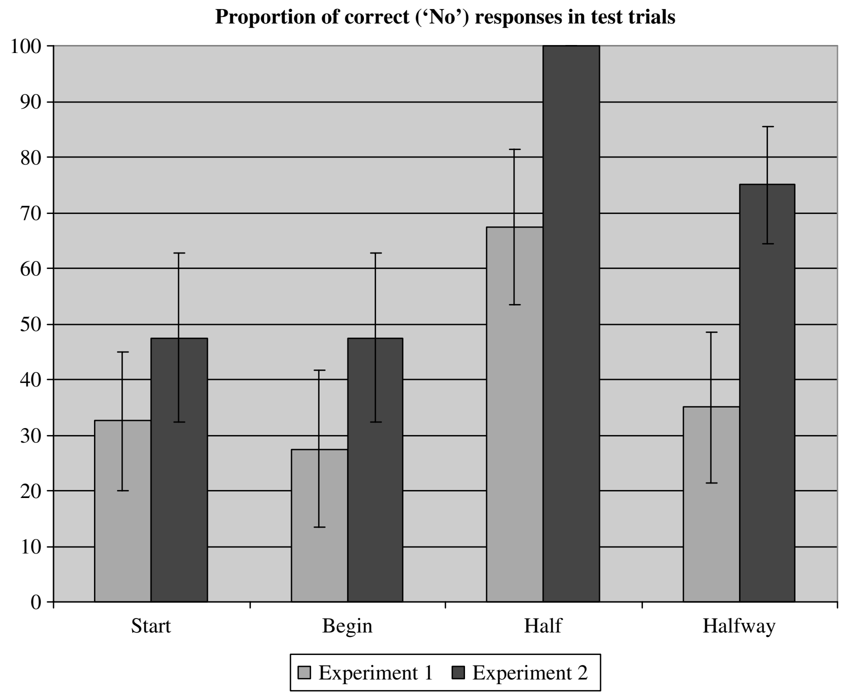

Fig. 1. Proportion of children's correct ('NO') responses in test trials.

conditions was not different from chance ($p = 0{\cdot}87$), unlike performance in the 'halfway' condition ($p = 0{\cdot}04$).[7]

A look at children's justifications for their refusal to award a prize in test items reveals that they always correctly made reference to a stronger scalar term (e.g. 'The horse didn't FINISH the school'). Finally, on the control items, children always gave correct answers in all conditions.

Recall that in Experiment 1, children's success rates for the four items tested were much lower (32·5% for 'start', 27·5% for 'begin', 67·5% for 'half' and 35% for 'halfway'). Children's answers in test trials from Experiments 1 and 2 were entered into a (2) experiment: 1, 2 × (4) expression: 'start', 'begin', 'half', 'halfway' ANOVA. The analysis revealed a significant main effect of experiment ($F(1, 72) = 8{\cdot}8$, $p = 0{\cdot}0039$) and expression ($F(3, 72) = 5{\cdot}5$, $p = 0{\cdot}0017$) but no interaction between the two ($F(3, 72) = 0{\cdot}4$, $p = 0{\cdot}75$) – see Figure 1.

[7] As in Experiment 1, we split children into Passers and Failers on the basis of their performance on test items. For 'start' and 'begin', we found 4 Passers and 5 Failers; for 'halfway', 7 passers and 1 Failer; for 'half', 10 Passers and no Failer.

DISCUSSION

Two important conclusions emerge from these findings. First, children in a naturalistic testing context are more likely to show competence with conversational implicatures than in a test situation which involves pragmatic judgment. As a comparison of results from Experiments 1 and 2 shows, the choice of task affects children's performance in computing scalar (non-completion) inferences from the use of aspectual terms such as 'start' or 'begin' and degree modifiers such as 'half' or 'halfway'. Methodologically, this result has implications for the design of studies which test pragmatic development, since it raises the possibility that certain widely used tasks may in fact mask early abilities for implicature-calculation.

A second conclusion which emerges from these findings is that, within the class of expressions which can be used to generate SIs, there are some interesting asymmetries. Five-year-olds in our study had difficulties with SIs triggered by the use of inchoative verbs ('start' and 'begin'), for which performance did not exceed chance levels; by contrast, their performance was very good with proportional modifiers, both in the nominal and in the verbal domain ('half', 'halfway'), even though not always perfect ('half' led to better performance than 'halfway').

Before discussing the reasons for these asymmetries, we need to rule out one possible explanation for children's behaviour in the above experiment. One might argue at this point that children around five may not yet fully know the semantics of aspectual and/or degree expressions. If this is the case, children's failures with 'start' and 'begin' may be attributed to lack of relevant semantic knowledge rather than inability to compute scalar pragmatics. A similar explanation could be offered for the 'half'–'halfway' discrepancy. At this stage some evidence is needed that Greek-speaking preschoolers have acquired the truth-conditional semantics of degree and aspectual terms. A third study was designed to demonstrate that children are sensitive to the lower-bounded interpretations of scalar terms in contexts where upper-bounding meanings are removed.

EXPERIMENT 3

METHOD

Participants

Participants in this experiment were a new group of 40 Greek-speaking children between the ages of 4;4 and 5;11 (mean age 5;6). Children participants were recruited from the same Athens daycare as in the other studies.

Materials and Procedure

Children were told that they would witness a drawing competition. Four animals, a lion, a bear, a giraffe and a horse, were given a sheet of paper and a pencil each and got ready to draw. Before the drawing started, children were informed of the terms of the competition, which were phrased in terms of one of the following statements:

(22) Whoever starts drawing a star gets a prize.
(23) Whoever begins drawing a star gets a prize.
(24) Whoever draws half a star gets a prize.
(25) Whoever draws a star halfway gets a prize.

The animals took turns in drawing but each one produced something different. The lion drew a star; the bear drew a circle; the giraffe drew half a star; and the horse drew nothing. Children, after being briefly reminded what the condition for winning a prize was, had to determine for each animal whether it should get a prize or not.

The aspectual/degree expressions in (22)–(25) appear in a DOWNWARD ENTAILING environment (i.e. in the scope of 'Whoever …', MGr *Opios*). An operator is downward entailing if it licenses inferences from a set to its subsets: (26) licenses the inference that whoever starts drawing a RED star gets a prize – hence 'whoever' is downward-entailing (and similarly for *Opios*). Downward-entailing environments have the property of inverting scales. For instance, even though (26b) entails (26a), (27a) entails (27b):

(26) a. He started drawing a star.
b. He finished drawing a star.
(27) a. Whoever starts drawing a star gets a prize.
b. Whoever finishes drawing a star gets a prize.

To illustrate the logic of scale reversal, consider what sorts of situations would satisfy the condition posed by the free relative clause in (27) – and hence warrant awarding a prize to a contestant:

(28) a. Whoever starts drawing a star gets a prize.
Situation 1: A draws 10% of a star.
Situation 2: B draws 20% of a star.
…
Situation n: Y draws 100% of a star.
b. Whoever finishes drawing a star gets a prize.
*Situation 1: A draws 10% of a star.
*Situation 2: B draws 20% of a star.
…
Situation n: Y draws 100% of a star.

https://doi.org/10.1017/S0305000906007550 Published online by Cambridge University Press

Clearly, the set of situations in (28b) is a subset of the set of situations in (28a) – hence the entailment pattern. Notice now that a weak scalar element in the scope of a downward entailing operator no longer invites scalar inferences, since there is no stronger scalar alternative the speaker could have used but did not. In other words, in downward-entailing contexts, scalar implicatures are suspended and scalar expressions receive lower-bounded ('at least') interpretations (Chierchia, 2004).

In our experimental setting, the clause introduced by 'whoever' in (22)–(25) places a minimal requirement on the contestants (drawing part/half of a star) and is compatible with actions which go beyond that (completion of the star). The prediction then is that, if children know the semantics of degree/aspectual expressions, even if they are not sensitive to their pragmatic (SI-triggering) properties, they should award a prize to both the animal which produced half of the star and to the animal which drew the whole star. The experimental design explicitly allowed for the possibility that the drawing competition would have more than one winner: many prizes (coins) were available in the scene and were ready to be distributed, if needed.

Children were randomly assigned to one of four groups depending on the kind of scalar expression used ('start', 'begin', 'half', 'halfway'). Children's ages ranged between 5;0 and 5;9 for the 'start' group (mean: 5;6), 4;11 and 5;11 for the 'begin' group (mean: 5;6), 4;4–5;11 for the 'half' group (mean: 5;6) and 5;0–5;9 for the 'halfway' group (mean: 5;6). Within each condition, a single trial was administered to participants.

RESULTS

Overall, 85% of the children gave errorless responses, i.e. (a) they awarded a prize to both the animal which drew a star and to the animal which drew half of the star, and (b) they didn't give a prize to anyone else. Errorless performance was observed on 80% of the responses in the *arxizo* ('start') group, 90% of the responses in the *ksekino* ('begin') group, 80% of the responses in the *miso* ('half') group and 90% of the responses for the *mexri ti mesi* ('halfway') group.

DISCUSSION

Results from the third experiment suggest that children around the age of five can assign the correct semantic content to both aspectual verbs such as 'start' and 'begin' and to degree modifiers such as 'half' and 'halfway'. Specifically, children can give a lower-bounded interpretation of these terms in environments in which the upper-bounded (scalar) implicatures typically associated with these expressions are cancelled.

These data are consistent with experimental results from English which suggest that children at age five (though not earlier) have mastered the semantics of degree terms. Wagner (2002) tested 2-, 4-, and 5-year-olds' understanding of completion/non-completion expressions (*all done, completely* vs. *partly done, in the middle of*). She used a sentence-to-picture matching task, in which children had to match a pair of sentences to two pictures, one of a completed and another of an incomplete event. What she found was that even the youngest children were successful with completion terms; with non-completion expressions, 5-year-olds were successful (even though the two younger age groups did not score above chance).

The findings from this task enable us to reject lack of semantic knowledge of aspectual/degree expressions as a possible explanation of children's pragmatic failures with such expressions in our previous Experiments. Specifically, insensitivity to the scalar pragmatics of 'start'/'begin' cannot be attributed to poor grasp of the semantic content of aspectual verbs. Relatedly, the fact that children are better at detecting the scalar inferences generated by 'half' than those triggered by 'halfway' is probably not due to lack of knowledge of the semantics of the latter. The issue of why children might generally fail with the SIs from inchoative verbs but succeed with the scalar inferences from the use of 'halfway' and especially 'half' is an important one, and we return to it in the General Discussion.

GENERAL DISCUSSION

Do children compute scalar inferences? And how do they discover the intricate division of labour between the lexical-semantic content of scalar expressions and the pragmatic inferences which these expressions typically give rise to? This paper was an experimental investigation into these two questions focusing on young children's interpretation of aspectual/degree expressions – an area where scalar semantics and pragmatics make tightly connected contributions to meaning.

Our studies were conducted with native speakers of Modern Greek and used four scalar expressions: *arxizo* ('start'), *ksekino* ('begin'), *miso* ('half') and *mexri ti mesi* ('until the midpoint/halfway'). These expressions typically give rise to scalar implicatures: for instance, the utterance *John started to paint the fence/painted the fence halfway* gives rise to the inference that John didn't complete the fence-painting. Our experiments asked whether preschoolers are able to compute such non-completion inferences; they also asked whether knowledge of the (implicature-free) semantics of these expressions surfaces in environments in which conversational implicatures are suspended.

The overall conclusion from our experiments is that scalar inferences, even though regularly computed when appropriate during adult language

https://doi.org/10.1017/S0305000906007550 Published online by Cambridge University Press

comprehension, are not reliably derived by otherwise linguistically savvy young children. Specifically, it was found that, even though 5-year-olds show some evidence of knowing the semantics of aspectual verbs and degree modifiers in implicature-free contexts (cf. Experiment 3), their ability to draw scalar inferences from these terms is not consistent. These findings extend and confirm previous results on the comprehension of scalar inference by children and adults (Noveck, 2001; Chierchia *et al.*, 2001; Gualmini *et al.*, 2001; Papafragou & Musolino, 2003, and others).

Perhaps the most striking finding of our experiments is that children's performance with SIs is not an all-or-nothing affair but depends on complex interactions of semantic and pragmatic factors. Children's success with SIs depends on the type of aspectual/degree expression: our two degree modifiers ('half', 'halfway') were, as a class, more successful SI-triggers than the two aspectual verbs ('start' and 'begin'), and 'half' was the most successful of all. Children's performance also crucially differs depending on the type of test used to measure pragmatic sophistication: a more naturalistic dialogue-like setting in which implicatures arise spontaneously (Experiment 2) leads to greater success than a pragmatic judgment task where children judge the felicity of utterances on the basis of their potential implicatures (Experiment 1). We now turn to these observations in more detail and consider their theoretical and methodological implications.

Asymmetries across scalars

One of the most interesting aspects of the present findings is the fact that children's ability to compute SIs does not surface uniformly across scalar expressions. Specifically, our experiments revealed that children often appear to ignore aspectual verbs when evaluating a statement (hence to judge *Daisy started to colour the star* to be acceptable in a context in which Daisy finished the star-colouring). By contrast, young hearers readily access non-completion inferences from the use of proportional modifiers such as 'halfway' and especially 'half' (e.g. *Daisy coloured the star halfway/half the star*). For children in our experiments, it turns out, well begun is not half done. These findings are unexpected but perhaps not mysterious, especially if we take a closer look at the profile of individual scalar expressions.

Notice that a proportional modifier such as 'half' seems to exclude stronger terms on the scale in a more definitive way, as shown in the following pair:

(29) a. She started/began colouring the star but she didn't finish colouring it.
b. ?She coloured half of the star but she didn't colour all of it.

In (29a), there is a contrast (marked by *but*) between the first and the second clause: this is evidence that the initiation of the star-colouring event does not immediately exclude its completion. By contrast, in (29b), asserting that half of the star was coloured seems to exclude termination of the event in a more definitive way. Intuitively, a possible explanation for the contrast is that, normally, an assertion with 'half' or 'halfway' presupposes a quite exact state of knowledge on the part of the speaker. This makes the scalar inference to 'not all' particularly strong and confirms the intuition that no completion of the action is imminent.

We want to argue that this difference is diagnostic of a deeper, SEMANTIC split within the scalar class. Recall that, on the standard neo-Gricean account, scalar elements such as numbers, quantifiers and aspectuals have a lower-bounded ('at least') semantics which is upper-bounded by a scalar inference. According to this account, *half* means 'at least half' and – through a SI – 'no more than half'. We now propose that, in contrast to aspectual verbs, *half* and *halfway* (and their Modern Greek counterparts) do not have lower-bounded denotations and that this semantic difference underlies the contrast in (29).[8]

To begin with, the claim that *half* is lower-bounded faces a number of difficulties on closer inspection. First, in linguistic incorporation, *half* clearly has an 'exact', not an 'at least' reading: a *half-painted fence* cannot be a fence that has been painted completely and a *half-hearted comment* is never whole-hearted. Second, *half* interpreted as a fraction in mathematical statements always has an 'exact' interpretation (cf. *one and a half centimeters*). Third, and perhaps most convincingly, on a range of linguistic diagnostics, *half* behaves differently from aspectuals and other lower-bounded scalars. For instance, interrogatives containing *half* and *start* return different values to negative answers: a negative response to the question *Did the girl start to colour the star?* is taken to mean that she coloured none of it. But if the question is *Did the girl colour half of the star?*, a negative answer means that she coloured either none or all of it.

---

[8] Even though we discuss 'half' and 'halfway' together, the two differ in our data. One possible reason for why 'half' is easier for children is that the scale <'all', 'half'> is much more accessible and more frequently used for contrastive effect than the scale <'completely', 'halfway'> (in English, *half* is massively more frequent than *halfway* – Francis & Kučera, 1982; this is probably true in Greek too, even though there are no word frequency sources we could check). Another possible reason stems from the syntactic differences between the two expressions: the PP *mexri ti mesi* ('halfway') modifies a measure provided by the direct object of the verb whereas the adjectival modifier *miso* ('half') modifies the object of the verb directly. This difference between object and event modification could be important: perhaps if a tower is built halfway, it is on its way to being built completely, but a tower that is half built does not indicate imminent completion.

Exactly the same range of arguments was originally used in the linguistic literature to argue that another class of scalar terms, namely numerals, should not be assigned lower-bounded semantics, as initially suggested (Horn, 1972) but should be treated differently from other scalar expressions (Sadock, 1984; Carston, 1990; Horn, 1992; Koenig, 1993; but cf. Levinson, 2000). Numbers, when incorporated, clearly have an 'exact', not an 'at least', reading: a four-sided figure has exactly four sides, and a four-seat plane has exactly four seats. Furthermore, numbers in mathematical statements are always assigned an 'exact' interpretation (*Four plus four makes eight*). Similarly, on the negative answer tests, numbers pattern like *half*, as the reader can quickly verify (see Horn, 2005).

As these facts demonstrate, 'half' turns out, perhaps unsurprisingly, to share semantic properties with number terms and to behave differently from other scalar expressions for which a lower-bounded ('at least') semantics is plausible. Furthermore, in both cases, there is a strong intuition that the upper-bounded interpretation ('half and no more', 'ten and no more') is not part of an implicature but contributes directly to the proposition expressed by the utterance. Within the semantic literature, it has been proposed that a way of handling the number facts is to attribute to numerals a semantically underspecified content which takes on a range of interpretations through a process of pragmatic enrichment (Carston, 1990). We now propose to generalize this approach to other scalar values such as *half*. One way of implementing this sort of proposal is to assume that *half/ten of the Xs* always means 'exactly half/ten of the Xs' but a contextual parameter needs to be fixed pragmatically so that reference is made to some situational sub-part of the world or the world as a whole (see Kratzer, 2003). If reference is restricted to a sub-part of the world, *half of the boys came* leaves open the possibility that more than half of the boys came (which is essentially the 'at least' reading). If reference includes the world as a whole, *half of the boys came* excludes higher values (the upper-bounded reading). Other things being equal, 'out of context' statements with *half* (and number terms) are expected to exhibit a preference for the upper-bounded reading (i.e. they should be interpreted with respect to the whole world), since such an interpretation yields an informationally stronger assertion (Kratzer, 2003). This nicely accounts for the fact that, in ordinary conversation, the preferred interpretation of *half* (and the numerals) is upper-bounded (i.e. *half* is interpreted as 'half and no more'). Finally, on this approach, the upper boundary on *half* and the numerals does not arise as part of an implicature but is a result of a pragmatic inference which contributes to the truth-conditional content of the utterance (and can itself be contextually removed).

The unified treatment of 'half' and number terms makes a couple of quite straightforward and testable predictions. First, numbers, on their preferred interpretation, should exclude higher values on the numerical scale in a more definitive way than 'vague' quantifiers such as *some*. This seems to be the case: the following contrast parallels (29):

(30) a. She coloured some of the stars but she didn't colour all of them. (in a set of 4)
b. ?She coloured 2 of the stars but she didn't colour all/4 of them. (in a set of 4)

Second, children should be very likely to resist lower-bounded ('at least') readings of the numerals, just as they do with *half*. This prediction is also borne out. Papafragou and Musolino (2003) compared Greek-speaking 5-year-olds' interpretation of cardinals vs. quantifiers in pragmatic judgment tasks. They found that children overwhelmingly rejected underinformative statements containing cardinals but not quantifiers such as *some*. For instance, in a scenario where a group of three horses had jumped over a fence, children almost always rejected the statement *Two of the horses jumped over the fence;* however, in the same conditions, children were much less likely to reject the statement *Some of the horses jumped over the fence*. The authors concluded that not all scales are born equal and that numbers, unlike other scalars, should be assigned an 'exact' semantics (for further experimentation and discussion of the 'exact' proposal see Musolino, 2004; Huang, Snedeker & Spelke, 2004; Papafragou & Schwarz, in press; Hurewitz, Papafragou, Gelman & Gleitman, in press). The present findings extend and confirm these results by suggesting that both numerals and 'half' resist lower-bounded interpretations and should be treated differently from regular, lower-bounded scalars.

The semantic reanalysis of *half* and the numerals has a number of further advantages. To begin with, it elegantly accounts for the fact that 'half', just like the cardinals, is particularly useful in approximations: 'half' and the number terms, although precisely definable, are frequently used loosely to refer to quantities which normally fall outside their denotation (*half the time; a hundred times*; cf. Dehaine & Mehler, 1992). From a developmental point of view, the proposed semantics for 'half' squares well with a series of experimental findings which show that early proportional reasoning in children makes heavy use of the 'half' boundary (Spinillo & Bryant, 1999; Singer-Freeman & Goswami, 2001). Furthermore, this semantic view offers natural links to theories of numerical cognition and its development, which standardly assume that cardinals (and 'half') are discrete (Gelman & Gallistel, 1978).

To summarize, a difference between scalar expressions which has been proposed for independent reasons within semantic theory turns out to have consequences for the process of language acquisition. Specifically, the present study taken together with other studies of scalars (e.g. Papafragou &

Musolino, 2003; Musolino, 2004) supports a principled and robust developmental asymmetry between 'exact' scalar values such as *half* and the numerals on the one hand, and lower-bounded scalars such as aspectuals (*start, begin*) or quantifiers (*some*) on the other. From a learning perspective, this asymmetry suggests that superficially similar scalar expressions may, in fact, be acquired and processed by children via different mechanisms. For instance, it is tempting to conclude from these findings that children are more sensitive to pragmatic inferences which directly contribute to the truth-conditional content of the utterance than to true scalar implicatures. Although still a matter for speculation, preliminary support for this conclusion comes from the fact that children have some success in performing other 'local' pragmatic inferences which contribute to the proposition expressed by the utterance, such as narrowing down referential domains (e.g. ***The children*** *are eating*) or performing bridging inferences (*A car is passing by*. ***The door*** *is green*; Avrutin & Coopmans, 2000). Interestingly, as Experiment 3 shows, children treat both scalar 'enrichments' and scalar implicatures as defeasible: for instance, in downward-entailing contexts, children correctly remove the upper-bounded interpretation of both aspectual verbs such as *start* and degree modifiers such as *half*.

A final point: the aspectual cases demonstrate the relevance of developmental data for the formulation of semantic theory. Taken together, recent data from the behaviour of scalars in child language offer supporting evidence for the proper semantic analysis of scalars within the adult lexicon – specifically, a split analysis which treats *half* (and the numerals) differently from regular lower-bounded scalar expressions. This is a clear demonstration of how the study of 'mere' performance can affect theories of competence – and the study of 'mere' children can inform theories of knowledge structure in adults.[9]

[9] The present findings also have implications for the acquisition of *grammatical* tense/aspect systems, where the role of pragmatic/scalar phenomena has not been recognized. Consider the distinction between imperfective (*John was drawing a star*) and perfective aspect (*John drew a star*). Several studies have found that children can match imperfective sentences to a picture of an incomplete event and perfective sentences to a picture of a complete event (Weist, Wysocka & Lyytinen, 1991; Weist, Lyytinen, Wysocka & Atanassova, 1997). This result has been considered as evidence for children's early knowledge of the correct semantics for aspectual morphology. However, the result is hard to interpret in *semantic* terms, since the 'non-completion' inference from imperfectives is not a semantic entailment but a SI (Levinson, 2000). Other evidence suggests that adults treat imperfectives as being about an incomplete event only when they are contrasted with perfectives; children, however, are not affected by such a contrast and match imperfectives to complete and incomplete events indiscriminately (Wagner, 2002). These adult data make sense if we consider 'non-completion' readings from imperfective sentences as context-dependent scalar inferences; moreover, the child data confirm the conclusion of the present paper that 5-year-olds have trouble with (at least some) non-completion SIs. Several puzzles remain in this area (for instance, one

Asymmetries across experimental tasks

One conclusion which emerges from the present data is that five-year-olds are sometimes able to assess informativeness expectations as conversations unfold and – at least in some cases – compute conversational (scalar) inferences when such expectations are not met. Nevertheless, their performance depends on the nature and specific requirements of the tasks used to measure pragmatic sophistication.

Recall that the logic behind scalar inference is that the hearer evaluates what the speaker has uttered against a set of ordered alternatives (a scale), given certain expectations about what the speaker intended to communicate. So far one way of testing this reasoning in children (and adults) has been to use pragmatic judgment tasks (i.e. tasks targeting felicity, appropriateness, etc. of a statement). These tasks typically involve two communicative episodes. In the first episode, a character utters a sentence containing a weak scalar expression (e.g. *The tiger started to build the tower* in a situation in which the tiger finished building it). In the second episode, the experimenter asks the child whether the character gave a good description of the event. In order to pass the task, the child has to (a) correctly interpret the character's statement; (b) correctly interpret the experimenter's question (e.g. grasp what counts as a 'good description' of the witnessed event).

Despite their usefulness as an initial tool for uncovering hearers' sensitivity to pragmatic inference, judgment tasks of this sort are a less-than-satisfactory measure of pragmatic competence for a variety of reasons. Recall that, according to the neo-Gricean picture, the driving force behind the computation of a SI in an exchange such as (31b) is the availability of the stronger alternative in (32):

(31) a. Experimenter: What happened?
     b. Puppet: The tiger started to build the tower.
(32) The tiger finished building the tower.

More precisely, following the schema sketched in the Introduction, a hearer who hears (31b) is entitled to reason as follows:

(33) i. *Start* is lower-bounded by its literal meaning (*start* = 'at least start').
     ii. There is a stronger statement than (31b), namely (32), such that the latter unilaterally entails the former but not vice versa.

would want to know how very young children in the Weist *et al*. studies succeeded with 'non-completion' readings of the imperfective if they lacked a proper understanding of scalar inference). A pragmatically informed approach to the acquisition of the grammar of tense/aspect seems to us a promising way of investigating such puzzles.

iii. Given quantity requirements, if the speaker knows the information in (32), and that it would be relevant to me to know this fact, it would be misleading for her to utter (31b).
iv. The speaker is assumed to be co-operative and observe the conversational maxims (including quantity).
v. Therefore, the reason the speaker chose not to express the stronger proposition in (32) is that the speaker didn't know for a fact that it was true.
vi. I conclude that, for all the speaker knows, (32) is false (=scalar implicature).

We know that children can compute step (i) (see Experiment 3) and can probably compute the stronger alternative in step (ii) from the visual scene. Notice, however, that children may still be unable to reach the conclusion in (v) because the intermediate steps in (iii) and (iv) are not warranted in this experimental setup. Step (iii) presupposes that the speaker knows that a stronger (more informative) statement would be relevant for the hearer: however, this assumption may not be justified given the fact that the character describing what happened in our version of the task is known to be incompetent (a 'silly puppet'). Similarly, step (iv) assumes that the speaker is co-operative and observes the conversational demands of informativeness and relevance; neither of these assumptions may hold given the puppet's history of inappropriate conversational behaviour. Unlike genuine cases of implicature-calculation, the character offering the weak scalar statement in this line of studies does not actually *intend* the statement to give rise to a scalar implicature (and expect the hearer to recover that implicature as part of what the speaker *meant*): rather the conversationally inept character simply happens to choose a statement which conveys a *potential*, but not *actual* (intended, communicated) implicature.

In the second part of the task, participants are asked to judge whether the puppet gave a good description of what happened in the experimental scenario. In order to answer this question, participants need to estimate the experimenter's expectations about what a good description would have been and judge how the statement produced by the puppet fares compared to other possible alternatives. For children, this process raises several interrelated difficulties. First, given the open-ended nature of the question, it is not clear what the experimenter considers informative and relevant for the purpose of the exchange. Inferentially reconstructing what the experimenter considers as a conversationally appropriate contribution essentially amounts to grasping the experimenter's goals in setting up the task. For some children, perhaps, even a mildly relevant and informative statement could be considered acceptable, especially since the range of alternative responses is vast and the criterion for choosing one over another

https://doi.org/10.1017/S0305000906007550 Published online by Cambridge University Press

unspecified. The fact that the puppet is conversationally challenged might make children even more likely to consider a less-than-optimal description of the story as acceptable.

There is another, more fundamental sense, in which the experimenter's question is indeterminate. Notice that weak scalar statements in pragmatic judgment tasks are designed to be true, albeit underinformative (e.g. *The tiger started to build the tower*). Furthermore, these statements carry a (potential) implicature which is itself false (*The tiger didn't finish building the tower*). Such tasks are implicitly based on the assumption that participants will flag true but pragmatically inappropriate statements, i.e. they will take felicity rather than truth to be the basis of their acceptance or rejection of the target sentence. However, it remains an open possibility that children may take the purpose of the task to be the evaluation of the truth of the target statements. Even though we tried to discourage this possibility in the present experiments (e.g. by introducing a training phase which showed that true but pragmatically infelicitous statements should be rejected), it is still conceivable that children accepted the puppet's statements simply because they were true. In any case, this kind of situation is removed from the demands of naturalistic conversations, where linguistically encoded meaning and implicated meaning do not normally compete against each other but are both considered to be part of what the speaker intends to communicate.

There is some evidence that the difficulty of reconstructing the experimenter's goals in pragmatic tasks also affects adults' performance. Noveck (2001) found that adults when presented with out-of-context underinformative statements such as *Some airplanes have wings* agree with the statements about half of the time. Similarly, Papafragou & Schwarz (in press) found that in 56% of cases adults considered a promise such as *I'm going to colour most of the star blue* to be satisfied if all of the star was coloured blue. Thus adults, who are otherwise able to compute SIs, in the absence of linguistic or extra-linguistic indications of the speaker's goals, are at chance accepting statements containing a weak scalar term (e.g. *some, most*) in circumstances which make the corresponding stronger alternative (e.g. *all*) true.[10] More generally, there is overwhelming evidence from both children and adults that difficulty in estimating the

[10] When the scalar statement is presented out of context, participants typically try to reconstruct a conversational scenario in order to decide how to interpret the statement. In such impoverished environments, participants may have even greater difficulty estimating the experimenter's goals than when participants are asked to judge the appropriateness of a real speaker's statement. This may explain why adult participants in those studies were more reluctant to treat SIs as part of what was communicated than adults who participated in the present experiments.

experimenter's goals when interpreting the test question interferes with performance in a variety of tasks (Gelman & Greeno, 1989; Sperber, Cara & Girotto, 1995).

The present paper explored a different method for testing for conversational implicature (Experiment 2). This method includes no explicit judgment component and no additional computational demands beyond ordinary processes of implicature calculation (as represented in the schema in (33)). Participants hear a conversation between the experimenter and another speaker who produces a scalar statement and indicate whether they took the speaker to convey a SI:

(34) a. Experimenter: Did you build the school?
b. Animal: I started building it.

This method reproduces naturalistic contexts in which implicatures are typically calculated. Unlike judgment tasks, here scalar inferences (e.g. *The speaker didn't finish building the school*) naturally emerge as part of what the speaker wanted to communicate. Importantly, the structure of the exchange provides a strong cue to what the experimenter considers informative and relevant: notice that the form of the question in (34a) provides a more constrained set of alternatives (i.e. possible answers) than the corresponding question in (31a). By restricting the set of possible alternatives to the scalar statement, the structure of the exchange facilitates the computation of a scalar inference (cf. step (33iii) above)). Under these conditions, children's awareness of scalar inference improves considerably.

Children's performance in Experiment 2 raises the possibility that children's insensitivity to SIs reported in previous studies might be exaggerated by experimental features and not purely caused by inability to compute implicatures in exchanges with real interlocutors (cf. also Noveck, 2001; Papafragou & Musolino, 2003). It also suggests that further manipulation of context and presentation conditions may yield better performance on the part of young learners (cf. especially Papafragou & Tantalou, 2004). It would be interesting to extend the present method to contexts in which informativeness expectations would be inferred rather than explicitly set up by a preceding question as in (34a). Another possibility would be to manipulate the cost of computing alternatives, for instance by comparing implicatures derived from high- vs. low-contrast lexical items. Finally, it would be of interest to manipulate the salience of such alternatives by changing the focus structure of the task.

Beyond their methodological implications, the present set of studies raises broader issues about the scope and limitations of children's pragmatic abilities. Despite improved performance in Experiment 2, children's pragmatic sophistication overall falls short of adult-like levels. What is responsible for such early failures? Which aspects of the utterance-interpretation device develop so as to produce the interpretive flexibility manifest in adult hearers? Based on the results reported here, we suspect that part of the answer lies with children's ability to access and integrate different premises in computing speaker meaning, including their ability to infer expectations of informativeness/relevance and evaluate a linguistic stimulus with respect to other possible alternatives that the speaker could have selected. This kind of explanation is consistent with other evidence which shows that young children take limited advantage of contextual information in interpreting ambiguous linguistic strings and may only slowly develop an ability to integrate lexical and contextual cues in dealing with referential communication (Trueswell *et al.*, 1999).

Even though the computation of conversational inference is still fragile in 5-year-olds, our studies show that some fundamental properties of such inferences seem to be present in young learners. Recall that conversational inferences are assumed to be removable in context, hence cancelable. They are also taken to survive across synonymous expressions which belong to the same grammatical class and register level, hence to be non-detachable. With respect to cancellability, we found evidence that children can contextually suspend a conversational inference which they can otherwise compute. For instance, children treat 'half' and 'halfway' as compatible with 'all' and 'completely' respectively in downward entailing environments (Experiment 3), even though these interpretations were strongly dispreferred in inference-supporting contexts (Experiment 2). With respect to non-detachability, the data show that children treat synonyms of the same grammatical class equally with respect to their SI-generating potential. Specifically, children have the same (low) success rates with the SIs from 'start' and 'begin', just as non-detachability would predict.[11] To the extent that they are computed, then, early implicatures are well-behaved. This is an indication that the basic inferential mechanisms responsible for calculating implicatures in young children share at least some of the properties of the adult processor.

---

[11] Other aspects of our findings raise some questions about the scope and predictive power of non-detachability. Notice, for instance, that close semantic relatives (even though not synonyms) such as 'half' and 'halfway' do not behave identically with respect to their SI-generating potential: even though non-detachability does not apply to these cases, this asymmetry in the periphery of non-detachability is puzzling. Interestingly, a similar asymmetry was noted by Noveck (2001), who reports that the modals *could* and *might* did not give rise to SIs in similar ways in adult comprehension (again, probably because the two are not true synonyms). These observations, though not directly challenging the core of phenomena which non-detachability was designed to capture, raise some questions about the domain of application of non-detachability as a condition on implicatures (see Chierchia & McConnell-Ginet, 1990 for further criticism).

FINAL THOUGHTS

Most approaches to word learning to date have not addressed the fact that the interpretation of words may be context-dependent or that words in use give rise to a host of conversational inferences. Furthermore, the literature on semantic development sometimes implicitly assumes that pragmatic ('contextually enriched') interpretations are acquired at a later stage than 'pure' semantic meaning – hence it is possible to study the acquisition of semantics independently from pragmatic development. The results on aspectuals reported in this paper show that, in several respects, this approach is neither viable nor attractive. First, pragmatics may not be so hard: as our experiments show, children compute pragmatic inferences under certain conditions. Second, semantic content is not transparent: the theorist interested in studying pragmatically 'unadorned' semantic meaning needs to carefully construct contexts in which such meaning surfaces (e.g. implicature-canceling environments). Third, and crucially, this approach misses important generalizations about natural-language meaning: especially in the case of scalars, semantic and pragmatic contributions to speaker meaning are tightly linked. These observations jointly converge on the conclusion that theories of the acquisition of word meaning cannot be divorced from the study of the stable and systematic inferences which words give rise to in context. In certain cases, the lexical semantics and pragmatics of child language can even be a source of evidence for the structure of the adult lexicon.

This conclusion suggests a somewhat different perspective on the word learning task. As we have already discussed, in order to acquire word meanings, learners must be equipped with the ability to compare cross-situational word use in order to construct, test and modify hypotheses about lexical semantics. We now suggest that, during this process, patterns of word usage, rather than being irrelevant aspects of the word learning environment which need to be filtered out, deliver different regular and systematic information about word meaning and the contextual inferences it can give rise to. From a theoretical perspective, the interesting question is how the child extracts semantic content out of variable contextual interpretations of lexical items (e.g. *start*) and how this content, in turn, serves as input to the interpretive/processing system. From a methodological perspective, the challenge is to create experimental techniques to be used with young children that explicitly seek to separate the contribution of semantics and pragmatics in the interpretation of natural language expressions.

REFERENCES

Avrutin, S. & Coopmans, P. (2000). Children who build bridges. *Proceedings from the Boston University Conference on Child Language Development* **24**, 80–91. Somerville, MA: Cascadilla Press.

https://doi.org/10.1017/S0305000906007550 Published online by Cambridge University Press

Carston, R. (1990). Quantity maxims and generalised implicature. *UCL Working Papers in Linguistics* **2**, 1–31. Reprinted in *Lingua* **96** (1995), 213–44.
Chierchia, G. (2004). Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. In A. Belleti (ed.), *Structures and beyond*. Oxford: Oxford University Press.
Chierchia, G., Crain, S., Guasti, M. T., Gualmini, A. & Meroni, L. (2001). The acquisition of disjunction: evidence for a grammatical view of scalar implicatures. *BUCLD* **25**, 157–68. Somerville, MA: Cascadilla Press.
Chierchia, G. & McConnell-Ginet, S. (1990). *Meaning and grammar*. Cambridge, MA: MIT Press.
Crain, S. & Thornton, R. (1998). *Investigations in universal grammar: a guide to experiments in the acquisition of syntax and semantics*. Cambridge, MA: MIT Press.
Dehaene, S. & Mehler, J. (1992). Crosslinguistic regularities in the frequency of number words. *Cognition* **44**, 43–74.
Dowty, D. (1979). *Word meaning and Mbntague grammar*. Dordrecht: Reidel.
Dowty, D. (1991). Thematic proto-roles and argument selection. *Language* **67**, 547–619.
Francis, W. N. & Kučera, H. (1982). *Frequency analysis of English usage: lexicon and grammar*. Boston: Houghton Mifflin Company.
Gelman, R. & Gallistel, R. (1978). *The child's understanding of number*. Cambridge, MA: Harvard University Press.
Gelman, R. & Greeno, J. (1989). On the nature of competence: principles for understanding a domain. In L. Resnick (ed.), *Knowing and learning: essays in honor of Robert Glaser*. Hillsdale, NJ: Erlbaum.
Grice, P. (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.
Gualmini, A., Crain, S., Meroni, L., Chierchia, G. & Guasti, M. T. (2001). At the semantics/pragmatics interface in child language. *SALT* **XI**. Ithaca, NY: Cornell University.
Hay, J., Kennedy, C. & Levin, B. (1999). Scalar structure underlies telicity in 'degree achievements'. *SALT* **IX**, 127–44.
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, UCLA.
Horn, L. (1992). The said and the unsaid. *SALT* **II**, 163–92. Dept. of Linguistics, Ohio State University.
Horn, L. (2005). The border wars: a neo-Gricean perspective. In K. von Heusinger & K. Turner (ed.), *Where semantics meets pragmatics*. Oxford: Elsevier.
Huang, Y.-T., Snedeker, J. & Spelke, E. (2004). What exactly do numbers mean? Paper presented at the 26th Annual Meeting of the Cognitive Science Society, Chicago.
Hurewitz, F., Papafragou, A., Gleitman, L. & Gelman, R. (in press). Asymmetries in the acquisition of numbers and quantifiers. *Language Learning and Development*.
Jackendoff, R. (1996). The proper treatment of measuring out, telicity, and perhaps even quantification in English. *Natural Language and Linguistic Theory* **14**, 305–54.
Kazanina, N. & Phillips, C. (2003). Temporal frames-of-reference in the development of aspect. Paper presented at the Workshop on the Acquisition of Aspect, Berlin, May.
Koenig, J.-P. (1993). Scalar predicates and negation: punctual semantics and interval interpretations. *CLS Proceedings, Parasession on Negation*, 140–55.
Kratzer, A. (2003). Scalar implicatures: are there any? Talk delivered at the Milan Workshop on Scalar Phenomena, Polarity and Implicature. University of Milan-Biccoca, 13–16 June.
Krifka, M. (1992). Thematic relations as links between nominal reference and temporal constitution. In I. Sag & A. Szabolsci (eds), *Lexical matters*. Stanford: CSLI.
Levinson, S. (2000). *Presumptive meanings*. Cambridge, MA: MIT Press.
Musolino, J. (2004). The semantics and acquisition of number words: integrating linguistic and developmental perspectives. *Cognition* **93**, 1–41.

https://doi.org/10.1017/S0305000906007550 Published online by Cambridge University Press

Noveck, I. (2001). When children are more logical than adults: experimental investigations of scalar implicature. *Cognition* **78**, 165–88.
Noveck, I. A., Chierchia, G., Chevaux, F., Guelminger, R. & Sylvestre, E. (2002). Linguistic-pragmatic factors in interpreting disjunctions. *Thinking and Reasoning* **8**, 297–326.
Papafragou, A. & Musolino, J. (2003). Scalar implicatures: experiments at the semantics–pragmatics interface. *Cognition* **86**, 253–82.
Papafragou, A. & Schwarz, N. (in press). 'Most' wanted. *Language Acquisition.*
Papafragou, A. & Tantalou, N. (2004). Children's computation of implicatures. *Language Acquisition* **12**, 71–82.
Pinker, S. (1994). How could a child use verb syntax to learn verb semantics? *Lingua* **92**, 377–410.
Sadock, J. (1984). Whither radical pragmatics? In D. Shiffrin (ed.), *Georgetown University Roundtable on Language and Linguistics*. Washington, DC: Georgetown University Press.
Shatz, M. (1980). Communication. In P. Mussen (ed.), *Handbook of child psychology, vol. 3: Cognitive development* (vol. eds. J. Flavell & E. Markman), 841–89. New York: Wiley.
Singer-Freeman, K. & Goswami, U. (2001). Does half a pizza equal half a box of chocolates? Proportional matching in an analogy task. *Cognitive Development* **16**, 811–29.
Smith, C. (1980). Quantifiers and question answering in young children. *Journal of Experimental Child Psychology* **30**, 191–205.
Smith, C. (1991). *The parameter of aspect*. Dordrecht: Kluwer.
Sperber, D., Cara, F. & Girotto, V. (1995). Relevance theory explains the selection task. *Cognition* **57**, 31–95.
Sperber, D. & Wilson, D. (1986). *Relevance: communication and cognition*. 2nd ed. 1995. Cambridge, MA: Harvard University Press.
Spinillo, A. & Bryant, P. (1991). Children's proportional judgments: the importance of 'half'. *Child Development* **62**, 427–40.
Trueswell, J. C., Sekerina, I., Hilland, N. M. & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition* **73**, 89–134.
Van Hout, A. (2002). Imperfect imperfectives. Paper presented at the Workshop on Syntax, Semantics and Acquisition of Aspect, University of Iowa, May 24–26.
Verkuyl, H. (1993). *A theory of aspectuality*. Cambridge: CUP.
Wagner, L. (2002). Understanding completion entailments in the absence of agency cues. *Journal of Child Language* **29**, 109–25.
Weist, R., Lyytinen, P., Wysocka, H. & Atanassova, M. (1997). The interaction of language and thought in children's language acquisition: a crosslinguistic study. *Journal of Child Language* **24**, 81–121.
Weist, R., Wysocka, H. & Lyytinen, P. (1991). A crosslinguistic perspective on the development of temporal systems. *Journal of Child Language* **18**, 67–92.