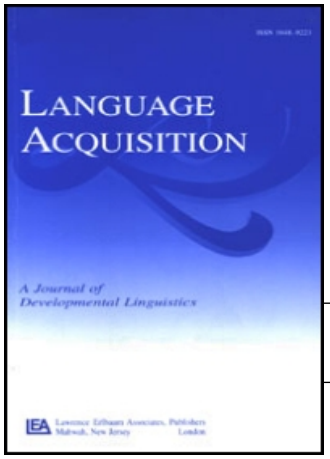

Language Acquisition
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t775653668

Most Wanted

Anna Papafragou [a]; Naomi Schwarz [b]
[a] University of Delaware.
[b] University of Pennsylvania.

Online Publication Date: 01 July 2006

To cite this Article: Papafragou, Anna and Schwarz, Naomi (2006) 'Most Wanted', Language Acquisition, 13:3, 207 — 251

To link to this article: DOI: 10.1207/s15327817la1303_3
URL: http://dx.doi.org/10.1207/s15327817la1303_3

LANGUAGE ACQUISITION, *13*(3), 207–251

Most Wanted

Anna Papafragou
University of Delaware

Naomi Schwarz
University of Pennsylvania

On the standard, neo-Gricean view, *most* is semantically lower bounded but may give rise to the meaning "not all" through scalar implicature (Horn (1972)). More recent proposals have claimed that *most* does not generate a scalar implicature but is semantically both lower and upper bounded (Ariel (2004; in press)). In this article, we investigate the interpretation of *most* experimentally to evaluate these competing semantic and pragmatic accounts. We focus on a comparison of *most* and *half* because, on the classical view, *half* and other exact determiners should admit bilateral interpretations more readily than the upward-oriented *most* (Horn (in press)); however, no such difference should exist if *most* is both lower and upper bounded. We find that (i) in nonbiasing contexts, adults are more likely to treat *most* as being compatible with *all* than *half*; (ii) a similar asymmetry emerges in children's interpretations of the two determiners; and (iii) adults adjust the higher boundary of the interpretation of *most* according to context-driven expectations. Taken together, these results support the classical, lower bounded, semantic analysis of *most* over recent revisions. Our findings also raise important issues about children's initial conjectures about scalar quantifiers and the development of the semantics–pragmatics interface.

1. INTRODUCTION

One measure of success for any linguistic theory is its ability to predict and explain data from both language comprehension in adults and language development in young learners. The importance of such predictive and explanatory power has been emphasized from the outset in modern linguistics (Chomsky (1965)). Conversely, several commentators have agreed that the way adults understand and process language can be an important source of evidence about the structure of underlying linguistic knowledge. Similarly, assuming basic *continuity* between the representational resources of children and adults (Pinker (1989)), the way lan-

Correspondence should be sent to Anna Papafragou, Department of Psychology, University of Delaware, Newark, DE 19716. E-mail: papafragou@psych.udel.edu

guage is acquired by children can be informative about properties of the adult grammar. On this integrated view, linguistic theorizing and psycholinguistic experimentation are expected to mutually inform and constrain each other.

In this article, we illustrate such an approach by focusing on a specific problem in linguistic theory: the semantics of *most*. *Most* and its relatives have been the topic of much debate in the literature (see Barwise and Cooper (1981), Horn (1989), Keenan and Stavi (1986), and the next section). The main goal of this article is to use experimental evidence from adults' and children's interpretation of *most* as a source of information for its correct characterization in adult grammar. Additionally, in the article, we aim at showing how specific theoretical analyses have implications for language acquisition and processing. We begin by surveying competing semantic and pragmatic accounts of *most* and their theoretical significance.

2. *MOST*: SEMANTICS AND PRAGMATICS

2.1. The Standard View

According to the classical, neo-Gricean position that goes back to Horn (1972), an utterance such as (1) typically gives rise to the inference in (2):

(1) Most gurus are wise.

(2) (For all the speaker knows) not all gurus are wise.[1]

To account for such inferences, it has been pointed out that quantifiers form a *scale*, that is, a set of alternates ordered in terms of informational strength:

(3) Quantificational scale: <*all, most, many, some*>

Informational strength is defined in terms of the asymmetric entailment relationships that hold between structurally equivalent propositions containing a member of the scale. For instance, *all* is informationally stronger than *most* within the quantificational scale in (3) because (4a) asymmetrically entails (4b) (and similarly for other scalars; Horn (1972)):

(4) a. All gurus are wise. →
b. Most gurus are wise. →

[1]In what follows, we omit the epistemic embedding for simplicity.

c. Many gurus are wise. →
d. Some gurus are wise.
(where → is a symbol of one-way entailment)

In the right contexts, the use of a weaker (lower ranked) alternate typically generates the inference that a stronger (higher ranked) member of the scale could not have been used. The reasons for this are tied to Gricean considerations of quantity/informativeness, according to which speakers should make their conversational contributions as informative as is required by the purposes of the exchange (Grice (1989)). In (1), for instance, assuming that the speaker is trying to be cooperative and will say as much as she or he truthfully can that is relevant to the exchange, the fact that the speaker chose a weaker term (e.g., *most*) from the quantificational scale gives the listener reason to think that the speaker is not in a position to offer an informationally stronger statement (using *all*)—presumably because such a statement is false. The inference in (2) is known as a *scalar implicature* (cf. Atlas (1984), Gazdar (1979), Grice (1989), Hirschberg (1985), Horn (1972), Levinson (2000); and for a different perspective Carston (1990; 1998), Sperber and Wilson (1986)).

More precisely, on this account, *most* has a lower bounded semantics ("more than half"), which can be rendered formally as follows (Keenan (1996)):

(5) Most (G)(W) = true iff $|G \cap W| > |G\text{-}W|$

For instance, in (1), *Most gurus are wise* is true in case the number of gurus who are wise is larger than the number of gurus who are not (i.e., if more than half of the gurus are wise). This literal meaning can be upper bounded by a scalar implicature ("more than half but not all"), which is computed through the following reasoning schema (adapted from Horn (1989, 214)):

i. The speaker has uttered (1).
ii. *Most* is lower bounded by its literal meaning (*most* = "more than half and possibly all").
iii. There is an equally simple but stronger statement than (1), that is, (4a), such that the latter unilaterally entails the former but not vice versa.
iv. Given quantity requirements, if the speaker knows that all gurus are wise and that it would be relevant to the hearer to know this fact, it would be misleading for the speaker to tell the hearer that most gurus are wise.
v. The speaker is assumed to be cooperative and observe the conversational maxims (including quantity).
vi. Therefore, the hearer infers that the reason the speaker chose not to express the stronger proposition in (4a) is that the speaker didn't know for a fact that it was true.

TABLE 1
Examples of Scalar Predicates

Scale Type	*Semantic Content*	*Communicated Content*
Quantifiers: <*all, some*>		
Pat ate some of the cake.	. . . at least some	. . . some but not all.
Modals: <*certain, possible*>		
It is possible she will come.	. . . at least possible	. . . possible but not certain.
Connectives: <*and, or*>		
John or Mary will win.	. . . at least one	. . . but not both.
Scalar adjectives: <*hot, warm*>		
It is warm.	. . . at least warm	. . . but not hot.
Adverbials: <*always, sometimes*>		
John sometimes sings.	. . . at least sometimes	. . . but not always.

vii. The hearer infers that, for all the speaker knows, (4a) is false, that is, not all gurus are wise (= (2)).

In sum, even though a simple statement of the form *S*(*most*) is compatible with *S*(*all*), it can be interpreted in conversation so as to exclude it. In that respect, *most* resembles not only other quantifiers (e.g., *some*) but also scalar expressions such as connectives (*or*), modals (*possible*), and so forth that are semantically unilateral but pragmatically bilateral (see Table 1; for discussion, see Horn (1972)). Notice that scalar implicatures, being a product of conversational rather than logical principles, are cancelable:

(6) Most gurus are wise—in fact, all of them are.

An important observation is that *most* is often used to convey a significant majority: (1) will be odd if just 51% of gurus are wise (Ariel (2004, 690), Horn (in press), McCawley (1981, 427)). Nevertheless, the upper boundary of the quantifier is subject to pragmatic adjustment. In other words, what counts as a significant majority will depend on various nonlinguistic factors including social or moral expectations and norms. For instance, an utterance of (7) invites the inference that an overwhelming majority of the population criticized the war, whereas in (8), a bare majority will suffice:[2]

(7) Most people criticized the war.

[2]The neo-Gricean view treats this inference as a conversational implicature (Horn (in press)). Another possibility is that the inference is a kind of pragmatic enrichment of literal meaning that contributes to truth-conditional content (on such enrichments, see Carston (2002), Sperber and Wilson (1986)).

(8) Most employees prefer this kind of retirement plan.

2.2. A Revised Semantics

Recently, the classical position on *most* has been challenged by Ariel (2003; 2004; in press). Ariel proposes that the lexical meaning of *most* is both lower and upper bounded and can be paraphrased as follows:

(9) Most Fs are G = "51% to 99% of (the) Fs are G"

On this proposal, *most* excludes reference to *all* by virtue of its lexical meaning. Ariel agrees that an utterance such as (1) is mainly used in contexts in which *not all* is true; however, Ariel does not consider (2) as a scalar implicature derivable from an utterance of (1) but a consequence of the fact that *most F* does not refer to *all F*. On the basis of naturally occurring data, Ariel argues that when speakers use *most*, they do not normally intend to implicate "not all" because such an implicature would defeat their conversational goals. For instance, in (7), because the speaker is presumably interested in making an antiwar statement, it would be hurtful to implicate that, in fact, not everyone criticized the war. Ariel concluded that "not all" should not be treated as part of a conversational implicature derived from ordinary uses of *most*: She proposes that *all* is simply outside the denotation of *most*—in other words, the semantic content of *most* is silent as to whether *all* is the case or not.[3]

On the revised account, the meaning of *most* coincides with its most frequent contextual interpretation as attested in natural conversations (Ariel (2004)): In other words, there is a direct connection between lexical content and attested contextual interpretations. As a further advantage of this account, Ariel (2004) suggests that the bilateral interpretation is a more natural concept to be encoded by a monomorphemic lexical item in natural language compared to the lower bounded ("at least") meaning proposed by the neo-Griceans.

Ariel (2003; 2004) offers two strands of experimental evidence for the revised analysis of *most*. First, she reports a series of questionnaire studies in which participants were asked to indicate acceptable interpretations of *most* using a variety of proportions (e.g., 49%, 50%, 75%, 99%, 100%, etc.). Results show that participants chose proportions over 50% as acceptable instances of *most*. Crucially, however, they strongly resisted choosing 100% as an acceptable instance of *most*, even when there was no obvious contextual block on the upper boundary. For instance, when presented with the utterance in (10), participants were unwilling to accept that the speaker *could* have had all of the students in the class in mind, even

[3]This is different from the proposal that *Most Fs are G* entails *Not all Fs are G* (for such a proposal, see Peterson (1979)). This last view cannot handle the fact that *most* is often treated as compatible with *all* (for further criticisms, see Horn (in press)).

though it is perfectly plausible that in a certain class, all of the students are born in the same year:

(10) Most of the students in this class were born in 1970.

Second, Ariel (2003; 2004) compared adults' interpretation of *most* versus *more than half*, the latter being an uncontroversial case of an expression with purely lower bounded semantics. Her results show that people are more willing to accept states of affairs in which *all* is the case as falling within the possible interpretations of utterances containing *more than half* than of utterances containing *most*. For instance, when presented with sentences such as (10), participants agreed only 6.25% of the time that the speaker could have considered it possible that all of the students were born in 1970. When *most* was substituted by *more than half,* the acceptance rate rose to 37.5% of responses. This finding is taken to support the conclusion that the exclusion of *all* is pragmatically derived via scalar inference in the case of *more than half* (hence, it is defeasible) but is part of semantic content in the case of *most* (hence, it is absolute).

2.3. *Most* and Other Scalars

The difference between the neo-Gricean and the revised account is brought out strongly by comparing *most* to numerals. As is well known, numerals form an entailment scale (< . . . , *three, two, one*>) and seem to give rise to the same scalar effects as quantifiers. For instance, an utterance such as (11) invites the scalar inference in (12):

(11) Three students came.

(12) No more than three students came.

As in the quantifier cases, the inference is cancelable:

(13) Three students came—in fact, four.

One way of capturing these similarities is to assign to numerals a lower bounded semantics that gets upper bounded by a scalar implicature (see Horn (1972) for such an analysis). On this neo-Gricean view, the semantic meaning of, say, *three* is "at least three"; this meaning, combined with a scalar implicature ("no more than three"), yields an "exactly three" interpretation. On closer inspection, however, numerals turn out to behave differently from other scalar expressions such as quantifiers (see Carston (1990), Geurts (1998), Horn (1992), Koenig (1991), Sadock (1984)). For instance, number terms readily accept an "at most"

reading that is unavailable for quantifiers such as *some*; it is hard to see how this reading can be accounted for on the standard, neo-Gricean analysis:

(14) She can eat three/?some ice-creams a day without putting on weight.

Second, numerals in incorporation have exact, not "at least" meanings (cf. *a two-door car, a four-sided figure*). Third, there are several ordinary uses of numerals where an "at least" semantics yields incorrect truth conditions. In (15), for instance, the "exact" interpretation for numerals is part of what is said and not the result of adding an upper bounding implicature to semantic interpretation:

(15) Two and two makes four.

More generally, it seems that the range of possible interpretations for number terms ("at least *n*," "at most *n*," "exactly *n*") contributes directly to the truth conditional content of the utterance (cf. (11), (14), and (15), respectively). For these and related reasons, it is now widely recognized that numerals do not submit happily to the standard neo-Gricean treatment of scalars (cf. Table 1) but need to be treated differently from other "inexact" members of the scalar class (e.g., quantifiers such as *some*). An attractive alternative to the neo-Gricean proposal is to assign to numerals a semantics that is underspecified between "exactly," "at most," and "at least" meanings (Carston (1990)). The semantic representation for *three,* on this view, can be rendered as follows:

(16) [*X* [three]]

On this analysis, the value of *X* is pragmatically enriched to yield the intended interpretation of the numeral ("at least three," "at most three," or "exactly three") depending on the specific context of use. This pragmatically enriched interpretation of the numeral affects the truth conditions of the utterance in which the number phrase belongs[4] (for alternative proposals, see Horn (in press), Koenig (1991); Levinson (2000) insists on a uniform neo-Gricean account of all scalars including numerals).

Consider now the relevance of the number cases for the semantic debates surrounding *most*. As the number cases demonstrate, scalar predicates are (at least sometimes) treated as logically compatible with a higher value on the scale they evoke (*three* is interpreted as "at least three"); nevertheless, this should not be con-

[4]On a recent (related) underspecification proposal, *three boys* always means "exactly three boys," but there is a kind of implicit parameter that needs to be fixed pragmatically so that reference is made to some situational subpart of the world or the world as a whole (Breheny (2005), Kratzer (2003)). If reference is restricted to a subpart of the world, *Three boys left* leaves open the possibility that more than three boys left (which is essentially the "at least" reading).

fused with the theoretical claim that the scalar meaning of these items specifies only a lower bound on this scale (a point first made by Koenig (1991)). Stated differently, the meaning of an expression may be compatible with a certain state of affairs without the expression encoding that state of affairs. Ariel (in press) explicitly builds on this distinction when she proposes to extend the reanalysis of the semantics of numerals to a reanalysis of the semantics of *most* along bilateral lines:

> Just like we are no longer forced to stipulate that *five* codes "at least five," just because it is not false to say *five* when "eight" is the case (. . .), so too the fact that we can use *most* when "all" is true should not force us to assume that the coded meaning of *most* covers "all."

The parallel between numerals and *most* is summarized in (17) and (18):

(17) Most students came to the demonstration.

(18) Ten students came to the demonstration.

On the revised analysis, (17) leaves it open whether in fact all students came to the demonstration, just as (18) leaves it open whether more than 10 students did so. In other words, both *most* and the cardinals are compatible with higher scalar values even though they are semantically uncommitted as to whether these higher scalar values hold or not.

Naturally, compatibility with higher values may be pragmatically suspended. For instance, Ariel (2004) recognizes that the use of bilateral *most* can still generate a scalar implicature ("not all") through the paradigmatic opposition with *all*, even though the quantifier encodes an upper boundary. The implicature will go through in cases in which *all* is relevant and expected (and occasionally, the "not all" inference can even affect the truth-evaluable content of the utterance).[5]

2.4. The Story of *Half*

The revised approach to numerals opens up the possibility that the unilateral semantic analysis for other scalars might also have to be abandoned. In this section, we propose that *half* is a good candidate for such a semantic reanalysis. Moreover, we suggest that a closer comparison of *most* and *half* can be used to evaluate the neo-Gricean and the bilateral views of the former.

There are good arguments for assuming that, like the cardinals, *half* lacks an "at least" semantics but has an underspecified content that acquires a range of interpretations through some process of pragmatic enrichment (see also Papafragou (2003; in press)). First, *half* can readily accept "at most" interpretations (e.g., *We*

[5]"Not all" inferences may also be computed by the hearer when warranted by the utterance even if the speaker did not intend to convey them (cf. (7)). Ariel (2004) calls these "uncooperative" inferences.

can fit half of the people in the car), which are unavailable for closely related scalar expressions such as *some*. Second, in incorporation, *half* clearly has an "exact" and not an "at least" reading: a *half-painted fence* cannot be a fence that has been painted completely. Third, *half* interpreted as a fraction (e.g., in mathematical statements) always has an "exact" interpretation; just like the cardinals, the "exact" interpretation seems to contribute directly to the truth-evaluable content of the proposition expressed rather than be part of an implicature. Finally, *half*, like the cardinals, is frequently used in approximations to denote meanings outside its denotation (e.g., *Half of my students showed up*). The conclusion seems to be that, just like the cardinals, *half* is different from other scalar expressions for which an "at least" semantics is plausible and may, in fact, have an underspecified content much like the one given for numbers in (16). On this analysis, the upper bounded interpretation of *half* ("half but not all") contributes to truth-conditional content and is not simply conversationally implicated.

Further linguistic diagnostics (originally proposed in Horn (in press) for cardinals) confirm that *half* patterns together with numerals and differently from *most*. For instance, simple negative answers to questions return different meanings depending on whether the question contains *most* or a cardinal/*half*:

(19) Did most of the participants show side effects?
  a. No, none did.
  b. ?No, all did.

(20) Did half of the participants show side effects?
  a. No, none did.
  b. No, all did.

(21) Did ten of the participants show side effects?
  a. No, eight did.
  b. No, twelve did.

The speaker in (19) can only negate the lower boundary of *most* because only this boundary is asserted. However, the answers in (20) and (21) leave it open whether the speaker is committed to a lower or a higher scalar value. Similar patterns emerge in embeddings:

(22) I'm surprised that most of the students solved the problem.

(23) I'm surprised that half of the students solved the problem.

(24) I'm surprised that ten of the students solved the problem.

The speaker in (22) is surprised that so many of the students solved the problem, not that some did not. However, in (23), the speaker could have expected that ei-

ther no student would solve it or that they all would; similarly, in (24), the expectation was that either fewer than 10 or more than 10 students would solve the problem.[6]

The more general point emerging from these diagnostics is that *most* differs from its scalar relatives (cardinals/*half*) in terms of its relation to higher scalar values. This conclusion is in line with the standard position on *most* according to which the quantifier is not upper bounded. However, the behavior of *most* is unexpected on the revised account. According to the bilateral argument we outlined in the previous section, other things being equal, people should have similar intuitions about the compatibility of both *most* and *half* with *all* (*most/half* are compatible with higher scalar values even though these values are not included in their denotation; see also Ariel (in press) and section 5). Nevertheless, this expectation seems to run contrary to the facts in (19) through (24). For instance, if *most* lexically encodes an upper boundary of 99%, why doesn't this boundary along with the lower boundary of 51% count as asserted, making (19b) felicitous? We return to these issues in the experimental part of the article in which we further explore hearers' intuitions about how *most* and *half* are interpreted and use these intuitions as a means of teasing apart the semantics and pragmatics of the quantifiers.

3. SCOPE AND EXPERIMENTAL PROSPECTUS

The differing views on *most* we have outlined continue an age-old debate in the logico-semantic literature between bilateral and unilateral approaches to quantifiers and other scalars. Bilateralists such as Hamilton (1860) sought to incorporate the upper bounded interpretation of quantifiers into their semantic meaning, whereas the unilateralist strand typically relegated upper bounding inferences to the extralogical domain (cf. De Morgan (1847); both cited in Horn (1989)):

> *Some*, if not otherwise qualified, means *some only*—this by presumption. (Hamilton (1860, 254))

> *Some*, in logic, means *one or more, it may be all*. He who says that *some are*, is not to be held to mean *the rest are not*. (De Morgan (1847, 56))

> In common conversation the affirmation of a part is meant to imply the denial of the remainder. Thus, by "some of the apples are ripe," it is always intended to signify that some are not ripe. (De Morgan (1847, 4))

[6]Notice also that cardinals, unlike *most,* allow scale reversal (Horn (1972, 44), Sadock (1984)):

(i) That golfer is capable of a round of 100 (and maybe even 90/*110).
(ii) She can counter most of the arguments (and maybe even *some/all).

Within this debate, a pertinent question has been whether all scalars should be treated uniformly or be subject to a mixed analysis whereby some of them would come out as bilateral and others as unilateral (see references in Horn (1989)). Perhaps unsurprisingly, it has been a natural move for unilateralists to generalize their approach to other scalars. De Morgan (1847), for instance, discussed how upper bounding inferences restrict the ordinary interpretations of a range of quantifiers in a distinctly neo-Gricean way:

> Common language makes a certain conventional approach to definiteness, which has been thrown away in works of logic. "Some" usually means a rather small fraction of the whole; a larger fraction would be expressed by "a good many"; and somewhat more than half by "most"; while a still larger proportion would be "a great majority" or "nearly all." (p. 58)

In the context of these and more recent discussions about the semantics and interpretation of quantifiers and scalar predicates (Chierchia (2004), Fox (2003), Kratzer (2003), Sauerland (2004)), the debate on the proper treatment of *most* raises two important issues. One issue concerns the architecture of the grammar of natural languages. Part of the appeal of the neo-Gricean position consists in offering a unified semantic–pragmatic account for scalar expressions that applies to a variety of items (quantifiers, logical connectives, modals, etc.) across all of the world's languages. Within each scale, individual expressions obey the same logic (i.e., one-sided semantics but two-sided pragmatic interpretation). If different scalar items turn out to behave in distinct ways (e.g., *most* and perhaps other quantifiers turn out to behave like cardinals), as the revised view expects, the appeal of the unified neo-Gricean program is diminished.

A second, related issue concerns language acquisition. The neo-Gricean picture simplifies the task of learning the semantic and pragmatic properties of individual members of the scalar class because the logic of the class generalizes to a variety of natural language items (see Table 1). However, if the scalar class is not homogeneous (as the revised position suggests), children cannot follow the same path in building conjectures about individual quantifiers but have to learn exceptions (e.g., *most*) on a case-by-case basis. Hence, the preceding semantic accounts differ not only in terms of what the child ultimately needs to learn about specific scalars (e.g., *most*) but also in terms of the kind of semantic generalizations available to learners during language acquisition.

There are further examples of such generalizations that emerge naturally from the neo-Gricean view but are harder to accommodate if scalars are internally split. Horn (1989) observed that certain types of quantifier meaning tend not to be grammaticalized in the world's languages. For instance, even though languages have quantifiers such as *none, some*, and *all*, they lack quantifiers such as **nall* (= "some but not all"), presumably because this meaning is typically communicated by *some* via scalar implicature. To the extent that this generalization becomes

available to children, it can greatly constrain the range of hypotheses about possible quantifier meanings. However, if we give up the neo-Gricean division of labor (lower bounded semantics plus upper bounding scalar implicatures) in favor of revised (bilateral) analyses for certain quantifiers, this generalization is lost.[7]

In this article, we use experimental data from both adults and children to adjudicate between the two different semantic and pragmatic accounts of *most*. Specifically, our experimental efforts aim at uncovering—and occasionally sharpening—the intuitions of native speakers about the preferred interpretations of *most* in different conversational contexts. These intuitions are then used as a source of evidence for the correct semantic and pragmatic characterization of the quantifier. In adopting this method, we concur with Barwise and Cooper (1981):

> Just which inferences involving "most" count as "logical" depends just where one draws the lines. . . . Our clues to the meaning of the strings of sounds represented by "most," and the inferential uses to which it may be put, are not determined by any logic writ in stone, but come only from the intuitions of native speakers. (p. 202)

A second goal of this work is to contribute to the discussion of the nature of scalar items, their processing, and acquisition. Are all scales born equal, and do asymmetries within the scalar class emerge in learners' early hypotheses about scalar meaning? Because little is known about the processing or acquisition of *most*, we expect these data to expand the empirical basis on which theories of scalars are built and evaluated. Finally, we hope that this sort of empirical investigation will further our understanding of how semantic information and pragmatic principles jointly contribute to natural language interpretation in both adults and children.

We structured the article as follows. We begin by asking how *most* is interpreted by adults in different contexts (Experiment 1). In Experiment 2, we compare *most* with *half*, a scalar value that seems to pattern alongside cardinals. In the experiment, we ask whether adults treat *most* as lower bounded (in accordance with the classical view) or, alternatively, whether they treat it as more similar to *half* in terms of admitting a bilateral interpretation. In Experiment 2, we also compare *most* to *more than half*, a clear case of an expression with purely lower bounded semantic content. In Experiment 3, we extend the comparison between adults' interpretation of *most* and *half* using different experimental materials and linguistic contexts. This experiment also includes three populations of young children; in it, we ask whether initial conjectures about the meaning of *most* treat it as a regular (lower bounded) scalar and whether there are developmental differences in how these conjectures are built. In the final section, we discuss our findings in

[7]Ariel (2004) considers the nonlexicalization facts a problem for the neo-Griceans because they pose an exception to the tendency of frequent conversational implicatures to become lexicalized. In Ariel's (2004) own account, *most* rarely gives rise to scalar implicatures, so there is no pressure for these meanings to be lexicalized.

the context of broader issues on the acquisition and processing of scalar terms and assess the implications of our experiments for the semantics and pragmatics of *most* and related scalars.

4. EXPERIMENT 1: ADULTS' INTERPRETATION OF *MOST*

This experiment is a first step toward establishing the correct semantic characterization of *most*. Furthermore, it provides evidence for context effects on the preferred interpretation of the quantifier.

4.1. Participants

Twenty-two adults participated in the experiment. They were mostly undergraduate students from the University of Pennsylvania.

4.2. Method

The task used in this experiment was a variant of the Truth Value Judgment Task (Crain and McKee (1985)). In this version, participants had to judge whether the outcome of the story (depicted in a picture) could be described by a statement containing *most*. Each experimental scenario included a "before" picture, in which one of the seven dwarfs or Snow White asserted that he or she was going to perform an action (e.g., color a star), and an "after" picture, in which the outcome of the action was pictured (e.g., a certain proportion of the star had been colored). We then asked participants to evaluate whether the agent had performed the action in a satisfactory way. Overall, participants had to judge whether the acceptable interpretations of *most* included the proportions 0%, 25%, 40%, 50%, 60%, and 75%. Preliminary piloting indicated that adults were able to compute these ratios accurately when inspecting the experimental stimuli.

The experiment had two conditions. The *neutral context* condition involved no special linguistic introduction. For example, in one trial, participants were shown Happy standing next to an empty fishbowl ("before" picture in Figure 1) and heard the following:

(25) Here's Happy. He said, "I'm going to fill most of this fishbowl with water."

Then the page was turned to reveal that the fishbowl had been filled up to 75% of its capacity ("after" picture in Figure 1) and the experimenter said

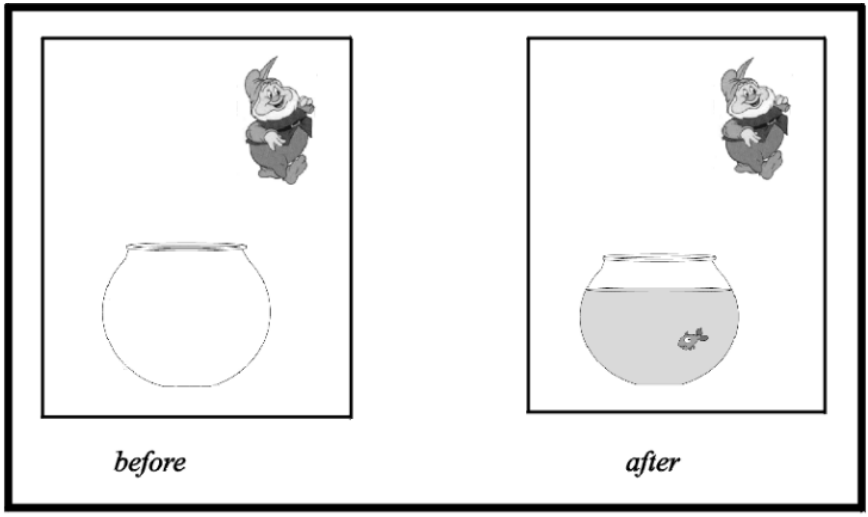

FIGURE 1 Experiment 1: A sample trial (75%).

(26) Look. Did he do okay?

In the *biasing context* condition, the character's statement was always preceded by another statement that raised contextual expectations about his or her performance (e.g., by emphasizing the need for the stated action, the character's exceptional prowess at the task, etc.):

(27) Here's Happy. He said, "My fish needs a lot of water. I'm going to fill most of this fishbowl with water."

Adults were again shown the story outcome (i.e., Happy with a fishbowl filled up to 75% of its capacity; Figure 1) and were asked whether the dwarf "did okay."

Each participant was assigned randomly to one of the two conditions in a between-subject design. Within each condition, each participant saw three trials with each proportion ($3 \times 6 = 18$ test trials in total). The pictures were designed so that the proportion was obvious—there was always a clearly defined "whole" (e.g., a container to be filled, a grouping of small objects to be manipulated) that underwent a salient change of state (e.g., in "filling" episodes, the filled portion was always spatially concentrated in a recognizable way). Likewise, the pictures were designed so that the proportions could be calculated without counting.

We also designed six control trials that were identical in format to the test trials within each condition but involved the quantifier *all*. The purpose of the control trials was to increase the number of expected affirmative answers overall. In one such trial, we presented participants with Snow White, who held up her hand:

(28) *neutral context condition:*
Here's Snow White. She said, "I'm going to color all of my fingernails red."

(29) *biasing context condition:*
Here's Snow White. She said, "I love red fingernails. I'm going to color all of my fingernails red."

The follow-up picture showed Snow White having colored all of her fingernails red, and we asked participants whether she did okay. We presented test and control trials to the participants in a pseudorandom order starting with a control item and intermixing the different kinds of proportions; order was counterbalanced across participants. For a full list of materials, see Appendix A.

Prior to the main testing phase, participants received three training trials that did not contain any quantifiers. They involved three dwarfs who were engaged in drawing circles. The first two training trials resembled test trials in nonbiasing contexts (*Here's dwarf X. He said, "I'm going to draw a circle." Look. Did he do okay?*). The first trial was meant to elicit a clear *yes* (a perfect circle was produced) and the second a clear *no* (nothing was drawn). The third training trial was modeled after the test trials in each condition and could receive one of two different answers. In the neutral condition, the dwarf simply said he was going to draw a circle. In the biasing condition, the dwarf announced that he was "the best at drawing shapes" and that he was going to draw a circle. In both cases, the dwarf went on to draw a squiggly circle. We expected that, in the neutral context condition, participants would accept that the dwarf did okay (i.e., fulfilled his promise), whereas in the biasing context condition, they would reject the outcome as incongruent with his statement. Most participants answered as expected; otherwise, they were given feedback by the experimenter.

Consider now the predictions made by different accounts of *most*. On the classical view, the neutral condition should allow for the full semantic range of interpretations of *most* (i.e., all proportions over 50%); however, the biasing context condition should narrow the legitimate interpretations of *most* to significant majorities only (i.e., 75% but not 60%). The revised view on *most* also shares the prediction that all ratios over 50% should be acceptable instances of *most* and, furthermore, that there should be context effects on the upper boundary of the quantifier.

4.3. Results

Results from the experiment are presented in Figure 2. (Participants performed without errors on control items; data from control trials are not included in Figure 2.) We conducted a repeated measures analysis of variance (ANOVA) using the number of *yes* responses as the dependent variable with trial type (0%, 25%, 40%,

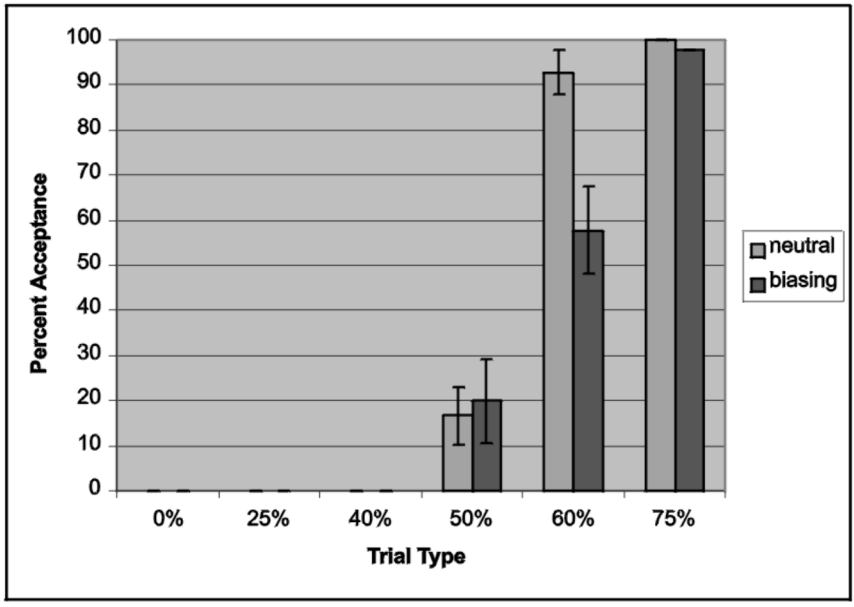

FIGURE 2 Experiment 1: Adults' interpretation of *most* in neutral and biasing contexts.

50%, 60%, 75%) as a within-subjects factor and condition (biasing, neutral context) as a between-subjects factor. The analysis showed no main effect of condition but a main effect of trial type, $F(5, 16) = 6.65$, $p < .01$, and an interaction between condition and trial type, $F(5, 16) = 0.60$, $p = .04$. Looking more closely at the results, we see that in the neutral context condition, adults correctly judged that 60% and 75% (but not 50%) are acceptable interpretations of *most* (proportion of acceptance: 92.86% and 100% vs. 16.67%, respectively). Pairwise comparisons showed that adults' responses did not change significantly between the neutral- and the biasing-context condition in the 50% trials (16.67% vs. 20% of acceptance rates, respectively) or in the 75% trials (100% vs. 97.78% of acceptance rates, respectively). Crucially, in the 60% trials, there was a significant difference between conditions (92.86% in the neutral vs. 57.78% in the biasing context condition; $p < .01$).

4.4. Discussion

Results from this experiment support the following two hypotheses: (i) *most* is semantically lower bounded, that is, it covers a range of proportions over 50%; and (ii) that lower boundary can be pragmatically adjusted. Together, these results ex-

tend prior psycholinguistic findings—mostly on vague quantifiers such as *few, some*, and *many*—which have demonstrated how contextual factors constrain the quantifiers' semantically permissible interpretations. For instance, Moxey and Sanford (1993, 29–30) reported experimental data showing that participants think that *some* corresponds to a larger number if it appears in the noun phrase *some people standing in front of the fire station* than if it appears in the phrase *some people standing in front of the cinema*, presumably because there is reason to expect a higher number of people in the first case. Similarly, estimates of numerical values are influenced by expectations of normal quantities. People might consider the sentence *Many tourists visited the zoo today*, uttered on a rainy summer day that also happens to be a national holiday, to be true if they compare the number of tourists with the normal number of visitors to the zoo on rainy days (not many), but false if they compare it with the normal number of visitors on national holidays (quite a few; Keenan and Stavi (1986, 258), cf. Drozd (2001, 357)).

Our findings can be accommodated by both the classical and the revised view of *most*. We next turn to the more controversial question of whether the semantically specified higher boundary of the quantifier should include or exclude *all*.

5. EXPERIMENT 2: A COMPARISON OF *MOST*, *HALF*, AND *MORE THAN HALF*

In Experiment 2, we directly tested the hypothesis that the lexical meaning of *most* does not include *all*. Recall that in Ariel's experiments it was found that "subjects consistently refused to confirm that a speaker could have referred to 100% when using *most*" (Ariel (2004, 676)). Here, we tested adults' interpretation of *most* using the range of ratios included in Experiment 1, adding contexts in which *all* (100%) is true.

In this experiment, we further compared adults' interpretation of *most* to their interpretation of *half*. Recall that, on the neo-Gricean view, *most* is treated differently from the cardinals (and *half*), whereas on the revised view, *most* and the cardinals (as well as *half*) should come out as similar to each other. In environments where *all* is the case, the neo-Gricean view predicts that *most* would be more likely to be tolerated than *half*, whereas the revised view, by taking both *most* and *half* to lexically exclude higher values, expects both to be largely unacceptable.

Finally, in the experiment, we compared the results on *most* to adults' interpretations of *more than half*. Recall that an important part of the original motivation for the bilateral analysis of *most* was provided by questionnaire studies that showed that *more than half* was far more likely to be considered compatible with *all* than *most* was (Ariel (2003; 2004)). This result was interpreted as a strong indication of semantic differences between the two expressions. More specifically, Ariel proposed that, even though *more than half* is lower bounded (and may be upper bounded by a scalar implicature of the familiar sort), *most* is both lower and

upper bounded (hence, it lexically excludes higher scalar values such as *all*). Here, we seek to test these same intuitions using our own experimental materials and procedure and explain their source.

5.1. Participants

Thirty adults participated in the experiment. All were undergraduate students at the University of Pennsylvania.

5.2. Method

In this experiment, participants had to judge whether the following proportions belong to the acceptable interpretations of *most, half,* and *more than half*: 0%, 25%, 40%, 50%, 60%, 75%, and 100%. The materials and design for this study were nearly identical to those of Experiment 1 (neutral context condition). For each trial, there was a "before" picture, in which Snow White or one of the seven dwarfs asserted that they were going to perform an action, and an "after" picture, in which we asked the participant to determine whether the action had been performed in a satisfactory way.

We assigned each participant randomly to one of three conditions (*most, half, more than half*) in a between-subject design. For instance, in one trial, participants saw Happy next to a star and heard one of the following:

(30) Here's Happy. He said, "I'm going to color most/half/more than half of the star blue."

Next, participants saw Happy next to a star completely colored blue and were asked:

(31) Look. Did he do okay?

Within each condition, each participant saw three trials with each proportion ($3 \times 7 = 21$ test trials in total). In addition to the 21 test trials, we inserted 4 control trials containing no quantifiers. The purpose of these trials was to increase the overall frequency of *no* responses. In 2 of the 4 control trials, the dwarf failed to do anything, and in the other 2, the dwarf completed a task other than the one promised. We presented test and control trials to the participants in a pseudo-random order, starting with a control item and intermixing the proportions in test items; we counterbalanced order across participants. For a full list of materials, see Appendix B.

As before, the main testing phase was preceded by a training phase. The overall format of the training trials was identical to that in Experiment 1 (neutral context condition) and involved three dwarfs engaged in drawing circles (*Here's dwarf X. He said, "I'm going to draw a circle." Look. Did he do okay?*). The first

training trial was meant to elicit a *yes* response (a perfect circle was produced), the second a *no* response (nothing was drawn), and the third a *yes* response (the dwarf drew a squiggly circle). Most participants answered as expected; otherwise, they were given feedback by the experimenter.

5.3. Results

Results from Experiment 2 are presented in Figures 3 through 5 (again, participants performed without errors on control items, and these data are omitted from the figures). We entered the number of *yes* responses into an ANOVA with quantifier (*most, half, more than half*) as a between-subjects variable and trial type (0%, 25%, 40%, 50%, 60%, 75%, 100%) as a within-subjects variable. The analysis revealed a main effect of quantifier, $F(2, 20) = 0.69$, $p = .001$; a main effect of trial type, $F(6, 20) = 14.43$, $p < .0001$; and a Quantifier × Trial Type interaction, $F(12, 20) = 12.80$, $p < .0001$.

Beginning with the data on *most*, we see that results confirm the findings from Experiment 1: In the absence of specific contextual expectations, adults accepted 60% or 75% but not proportions lower than or equal to 50% as legitimate interpretations of *most* (91.66% vs. 20% of acceptance rates, respectively; $p < .05$). Crucially, in the same circumstances, adults accepted *most* when *all* (100%) was true only 56.67% of the time (this value is not significantly different from chance; $p = .83$). A closer look reveals that adults had systematic preferences on the 100% trials: Four adults always said the dwarf did okay; another 1 gave this response in ⅔ of the 100% trials; the remaining 5 participants always refused to assert that the dwarf did okay.

Adults without exception accepted 50% trials as instances of *half* (100% of responses). They also accepted 60% trials (56.67% of responses) and occasionally 40% trials (36.67% of responses) probably because *half* can be used loosely in ap-

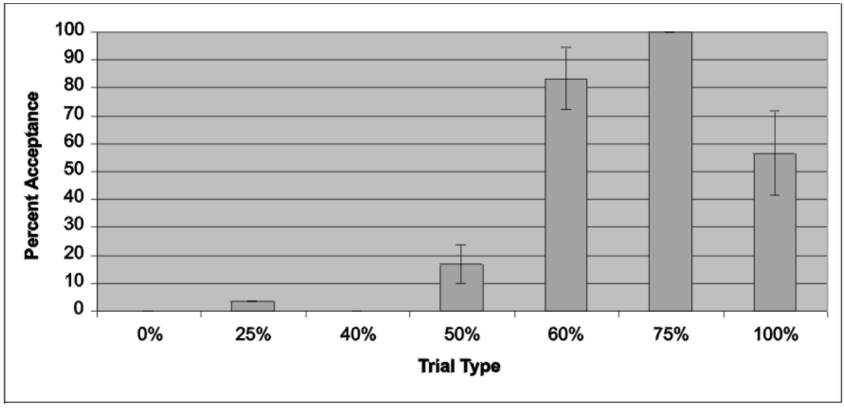

FIGURE 3 Experiment 2: Adults' interpretation of *most*.

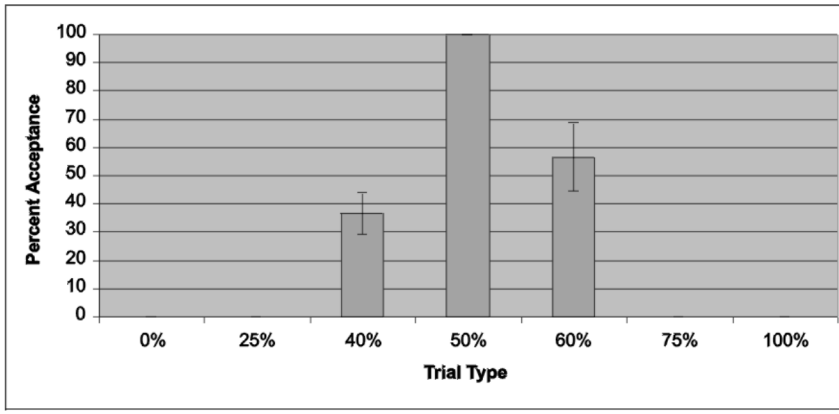

FIGURE 4 Experiment 2: Adults' interpretation of *half*.

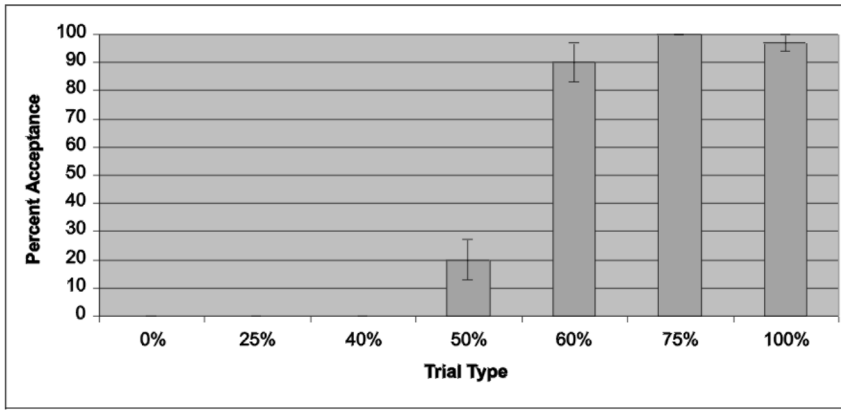

FIGURE 5 Experiment 2: Adults' interpretation of *more than half*.

proximations to refer to quantities around the 50% boundary (cf. Lasersohn (1999), Sperber and Wilson (1985–1986) on such uses). For our purposes, the most striking aspect of the data is that adults without exception excluded 100% from the range of acceptable interpretations of *half*. The difference in acceptance rates in the 100% trials between *most* and *half* (56.67% vs. 0% of responses, respectively) was statistically confirmed ($p = .0035$).

Finally, adults accepted proportions over 50% as acceptable instances of *more than half*, including 60% (90% of responses), 75% (100% of responses), and 100% (97% of responses). We entered the number of *yes* responses in the *most* and *more than half* conditions into a repeated measures ANOVA with quantifier (*most, more than half*) as a between-subjects variable and trial type as a within-subjects variable. The analysis revealed a main effect of trial type, $F(1, 11) = 20.56$, $p < .0001$, but no main effect of quantifier and no interaction between trial

type and quantifier. Pairwise comparisons between acceptance rates on individual trial types revealed that participants' interpretation of *most* and *more than half* yielded significantly different results only in the 100% trials ($p = .002$): In those trials, adults were split in their acceptance judgments for *most* but overwhelmingly accepted statements with *more than half.*

5.4. Discussion

5.4.1. Most *versus* half. Our findings offer evidence for the following conclusions: (i) *most* does not automatically exclude *all* by virtue of its lexical meaning (in fact, adults were split in their acceptance rates for *most* when *all* was true) and (ii) *half* is more likely to be interpreted as excluding the higher scalar value *all* than *most* is.

We look at the crucial 100% trials in more detail. On the neo-Gricean account, the utterance *I'm going to color most/half of the star blue* (cf. (30)) is infelicitous in case the dwarf turns out to have colored all of the star: If the dwarf's intention were to color all of the star, he should have said so. Given that *all*, if true, would have been relevant and expected, the fact that it was not used leaves open two possibilities. One possibility is that the dwarf did not know whether he was going to color all of the star. After all, dwarfs are small and fragile creatures and, as shown in the training trials of the task, often fail to perform simple everyday activities. It is therefore possible that the speaker could not anticipate the termination of the action he was planning and only wanted to commit to a partial completion. Another possibility is that the dwarf did not use *all* because he actually knew that he would not color all of the star. This might have been because he was unwilling or unable to perform the action in its entirety. On this interpretation, *most* and *half* invite a scalar inference, that is, the meanings they communicate are upper bounded. These interpretive possibilities are schematized in (32a,b):[8]

(32) a. ~ K (all)
Sp does not know whether he's going to color all of the star blue.
b. K ~ (all)
Sp knows that he's not going to color all of the star blue.

Which of these interpretations is more likely to have been intended by the use of *most* and *half* in the original statement? In the case of *most*, either interpretation is possible, given that the task includes no specific assumptions about what might affect the dwarf's performance. However, in the case of *half*, the two possibilities in (32a,b) are not equally likely: *half* is an "exact" scalar value, and by choosing

[8](32a) and (32b) correspond to Steps (vi) and (vii), respectively, in the computational schema for scalar implicatures (section 2.1). The two are inferentially related because (32b) is a stronger inference than (32a).

such a value, the speaker reveals a more precise state of knowledge than a vague (lower bounded) scalar could afford. This makes it more likely that (32b) is the case, that is, that the speaker has reasons to believe that "not all" rather than leaving it open whether "all" is true. Hence, the neo-Gricean view explains why the preferred interpretation of *half* in our task should be upper bounded; it also explains why lower bounded, upward oriented *most* should be more likely than *half* to tolerate "all."[9]

Our results argue against the bilateralist claim that people do not accept that "a speaker could have referred to 100% when using *most*" (Ariel (2004, 676)). Nevertheless, the bilateralist could still attempt to account for these data. For instance, partly in response to findings from this experiment, Ariel (in press) acknowledges the *half/most* asymmetry but attributes it to independent factors. Ariel points out that (underspecified) *half* is "punctual" and is separated by a long distance from the endpoint of the scale (*all*); by contrast, (bilateral) *most* is "wide range," and its distance from the endpoint of the scale is vanishingly short. Ariel (in press) concludes that these factors make *most* more likely to be judged compatible with *all* than *half* is. However, it is not clear how either of these factors predicts the observed difference. Consider the explicitly wide-range expression *between 51% and 99%*: The very fact that the speaker has chosen to limit the range so as to exclude *all* makes it less likely that *all* would be acceptable. A similar point applies to *nearly all*. These intuitions can be further bolstered by submitting these expressions to linguistic tests. As already discussed, *half* patterns differently from *most* with respect to its upper boundary on a range of distributional diagnostics (cf. (19)–(24)). For instance, in (33), the speaker expresses surprise at the fact that over 51% of the people stayed, not at the fact that less than 100% did; however, if *most* is replaced by *half* (cf. (34)), the speaker can be surprised by either how few or how many people actually stayed (i.e., both the higher and the lower boundary are asserted). Interestingly, the bilateral paraphrases of *most* in (35) and (36) pattern together with *half*, not *most*:

(33) I'm surprised that most of the people stayed.

(34) I'm surprised that half of the people stayed.

(35) I'm surprised that between 51% and 99% of the people stayed.

[9]Given that both semantic and pragmatic considerations affected participants' judgments in our task, we expected that changes in the presentation conditions and/or task demands should affect the preferred interpretation of *most/half*. For instance, in our study, quantifiers were embedded in utterances that described future actions and hence carried the illocutionary force of a prediction. This made it more likely that not all aspects of the action could have been foreseen by the agent and therefore that the options in (32) remained open. If the speaker knew that "all" was true, the acceptability of *most* in the 100% trials would decrease. This leads to the following prediction: If participants were presented with a picture of a completely colored star and heard the dwarf say *I colored most of the star blue*, they would be much less likely to consider the utterance felicitous.

(36) I'm surprised that nearly all of the people stayed.

These facts, together with our experimental findings, cohere with the traditional, lower bounded view but are harder to explain on the revised view of *most*.

Turning to the high acceptance of *most* in the 100% trials, the bilateralist could argue that at least some of the people in our experiment treated the question "Did he do OK?" as being about compatibility with the state of affairs denoted by the original proposition rather than about the proposition's meaning (Ariel (2004, 689))—thus accepting *most* even when *all* was the case. In her own studies, Ariel (2004, 690) suggests that participants' responses to *most* differ according to whether they are attending to truth compatibility with higher scalar values rather than semantic meaning—in the first case, they are very likely to accept *most* when *all* is in fact true. For instance, when presented with the sentence *The teacher already knows most of the students* and asked whether the speaker could have meant "51% to 100% of the students" (a semantic question, according to Ariel), participants refused; but when asked whether the sentence is true in case the teacher already knows all of the students (a compatibility question for Ariel), the majority of respondents said "yes."

Several other studies have confirmed that people's willingness to accept a weak scalar statement when a stronger one would have been true varies depending on how task demands are perceived. When asked whether they agreed with true but underinformative statements such as *Some elephants have trunks*, adults were split in their responses (Noveck (2001)). However, if asked whether statements of the form *Some Xs Y* were good descriptions of scenes that made *All Xs Y* true, adults consistently answered negatively (Papafragou and Musolino (2003); see also Papafragou (in press)). The question is whether these tests, together with Ariel's (2004) tasks and our experiment, target truth incompatibility (as Ariel (2004) holds) or scalar implicature (as the standard analysis of these results maintains). In either case, the variability in participants' naïve judgments about whether a weaker scalar is appropriate/true is explained in pragmatic terms and is hence expected to be context dependent; the two approaches differ in what they consider the source of this variability.

One problem with the bilateralist interpretation of the set of experimental findings on scalars is that it assumes that at least some of the tasks that have been used in the literature have targeted semantic meaning directly. In Ariel's (2004) tasks, by asking participants to indicate whether a speaker could (in principle) have had an "all" interpretation in mind when using *most*, she wanted to uncover what participants take the semantic meaning of *most* to be. However, it is well known that people's intuitions can only function as an indirect means of uncovering semantically encoded content. Several experimenters have already found that naïve participants' intuitions about meaning cannot reliably distinguish between sentence meaning (grammatically determined semantic value) and speaker meaning (semantic value plus implicatures; see Gibbs and Moise (1997) for a pertinent dem-

onstration and Koenig (2005) for discussion). In fact, most of the meaning tasks that have been used in the literature (including our own) cannot avoid some degree of pragmatic intrusion. Even if hearers are encouraged to consider all possible interpretations the speaker could have had in mind when uttering a statement (i.e., all the range of possible semantic options), it is very likely that the task will be contaminated by considerations of what the speaker could have intended to convey (i.e., pragmatics).

If we are correct, the difference between our own results and Ariel's can be explained as follows. Recall that one of Ariel's (2004) tasks involved selecting what a speaker could have had in mind when uttering a statement such as *The teacher already knows most of the students*. Appropriate paraphrases of *most* were provided (e.g., "51% to 100%," "80% to 95%," "51% to 75%"). However, a paraphrase task invites explicit comparison with linguistic alternatives (including *all*) while leaving the issue of how much the speaker knows about the teacher and the class wide open. Under those circumstances, participants assumed that if the speaker had intended to convey "all," he or she would have said so—hence, the majority of them rejected the paraphrase that best captures the standard semantics for *most* ("51–100%") and instead chose the pragmatically preferred interpretation for *most* ("80–95%"). In our design, we presented various utterances containing *most* in a prediction mode, and we later compared them to states of affairs that made *all* true. Under those circumstances, people were much more likely to say that the utterances came out true. Similarly, in Ariel's (2004) data, people overwhelmingly accepted that the statement *The teacher already knows most of the students* was true if the teacher turned out to know all of the students.

We conclude that the exclusion of higher scalar values by uses of *most* is not due to a lexically asserted upper boundary. To bolster this conclusion, in Experiment 3, we presented participants with a modified version of this task that explicitly targeted communicated meaning and therefore entirely removed the possibility of "compatibility" responses.

5.4.2. Most *versus* more than half. Finally, we turn to a particularly interesting aspect of our findings: *Most* and *more than half* differ in acceptability when *all* is the case. These findings support Ariel's (2003; 2004) empirical observation.[10] Nevertheless, this difference need not be captured in the semantics of the two expressions as Ariel (2003; 2004) proposed. We argue, following Horn (in press), that the discrepancy between *most* and *more than half* follows from inde-

[10]As with *most*, there were some significant differences between Ariel's (2003; 2004) data and our own. Specifically, Ariel's participants accepted *more than half* about one third of the time in 100% cases; by contrast, our participants accepted it in the overwhelming majority of 100% trials. The fact that *all* interpretations were overall more strongly resisted in Ariel's paraphrase task confirms the conclusion that her task invited the computation of scalar inferences more than the test we used here (cf. the discussion of *most* in the previous section).

pendent facts about the way the two expressions are typically used and is therefore perfectly compatible with a classical, unilateral approach to *most*.

Notice that even though both expressions impose a semantic lower bound, *most* is shorter and less marked than *more than half*; it is also closer in terms of lexicalization to *all* than *more than half* is. A speaker who chooses to use *more than half* rather than the shorter *most* must have a reason for doing so. As Horn (in press) suggested:

> One motive for using *more than half* may be precisely its compositional structure, which focuses on whether the proportion in question is less than, equal to, or greater than half of the set under consideration. This is likely to arise in situations involving near majorities or bare majorities. At the same time, if all that is relevant is whether a majority obtains, the presence or absence of totality is not necessarily relevant. These two factors explain why *more than half of the Fs are G* will be acceptable in many contexts in which just either 51% or 100% of the Fs are known to be G. But under these same two conditions, especially the latter, it is harder to imagine an acceptable use of *most Fs are G*, given that the only relevant choice will be that between *most* and its informationally weaker and stronger alternatives (in particular, *all*).

As Horn went on to explain, this does not make *most Fs are G* false in contexts in which *all Fs are G* is true, but such a sentence will be not only true but appropriately uttered as long as the speaker does not know that the *all* sentence is true when it is uttered.

These observations allow us to explain why, in our studies, a statement such as (37) excluded completion of the action more robustly than a statement such as (38):

(37) I'm going to color most of the star blue.

(38) I'm going to color more than half of the star blue.

In the *most* case, there was a clear, equally lexicalized strong alternative (*all*) that the speaker could have used if he intended his action to be carried through to completion; the fact that he didn't use *all* made the hearer hypothesize that either (i) the speaker did not know whether he would end up coloring all of the star (no scalar implicature; *most* judged compatible with *all*) or (ii) the speaker knew that he would stop short of coloring the whole star (scalar implicature generated; *most* judged incompatible with *all*). As we saw, adults are split between these two interpretations. In the *more than half* case, the choice of phrase makes the hearer consider whether the speaker will reach (and exceed) the *half* boundary: Whether the whole star will be colored is irrelevant. As a result, hearers are now willing to accept situations in which all of the star is colored as falling within what is communicated by (38).

To summarize, experimentally attested asymmetries in the acceptability of *most* and *more than half* can be explained on the basis of their formal (compositional) differences even though both expressions are true under the same set of circumstances. Therefore, the results of this experiment give us no reason to abandon the classic, lower bounded semantic analysis of *most* in favor of a bilateral approach.

6. EXPERIMENT 3: *MOST* VERSUS *HALF* IN COUNT CONTEXTS

In Experiment 3, we replicated the previous experiment with two methodological changes. First, we switched to quantifier use in count rather than mass contexts; specifically, we introduced an array of six objects and an agent (a dwarf) who says he will perform an action affecting half or most of them. Second, after the action is performed on a subset of the items, we asked whether the dwarf "did what he said." Even though this question does not distinguish between semantic and pragmatic contributions to "what was said" (cf. Gibbs and Moise (1997) and earlier discussion), it nevertheless excludes outcomes that are simply compatible with what the speaker said. Together with the use of count contexts, we hypothesized that this novel manipulation would focus participants' attention to what the speaker meant by the use of *most/half* and would discourage responses that take into consideration factors other than speaker meaning (e.g., the ease of approximation for mass quantities, whether an outcome is compatible with what the speaker had in mind, etc.).

In addition to adult participants, in the experiment, we also tested children's interpretation of *most* and *half*. Our main objective was to determine whether the difference between the two quantifiers in the "all" cases that was observed in the adult comprehension data would resurface with young children. More generally, we wanted to generate some data about the process of learning the semantics of the two quantifiers.

6.1. Participants

Twenty adults participated in the experiment. They were all undergraduate students at the University of Pennsylvania. Three age groups of children also participated. The younger group included 20 children ranging from 3;5 to 5;6 (*M* age = 4;5), the middle group included 18 children ranging from 5;11 to 8;3 (*M* age = 7;3), and the older group included 18 children who ranged from 8;5 to 11;11 (*M* age = 10;1). We recruited children from day cares in Philadelphia and in Newark, Delaware.

6.2. Method

Materials and procedure were similar to those of previous experiments. There were 24 trials, with 21 test trials and 3 controls. Each trial consisted of a "before" picture, in which Snow White or one of the seven dwarfs said they were going to complete an action, and an "after" picture, for which participants were asked to evaluate whether the character had completed the action stated. The test sentences always took the form of "I'm going to X most/half of these Y (with Z)." For example, were shown participants a picture of Grumpy next to six candles (see Figure 6) and heard the following:

(39) Here's Grumpy. He said, "I'm going to light most/half of these candles."
(The page was turned to reveal the outcome: Grumpy stands next to six lit candles.)
Look! Did he do what he said?

We tested participants on seven different ratios—0/6, 1/6, 2/6, 3/6, 4/6, 5/6, and 6/6—with three test trials for each ratio. All "before" pictures included six distinct countable objects arranged in two horizontal rows of three. These objects were acted on (e.g., consumed, transformed, or filled with new objects). Whenever an object disappeared, something in the picture served as a visual reminder of how many objects there were originally (e.g., one of the "before" pictures contained six pies on individual plates; in the "after" picture, some of the plates were

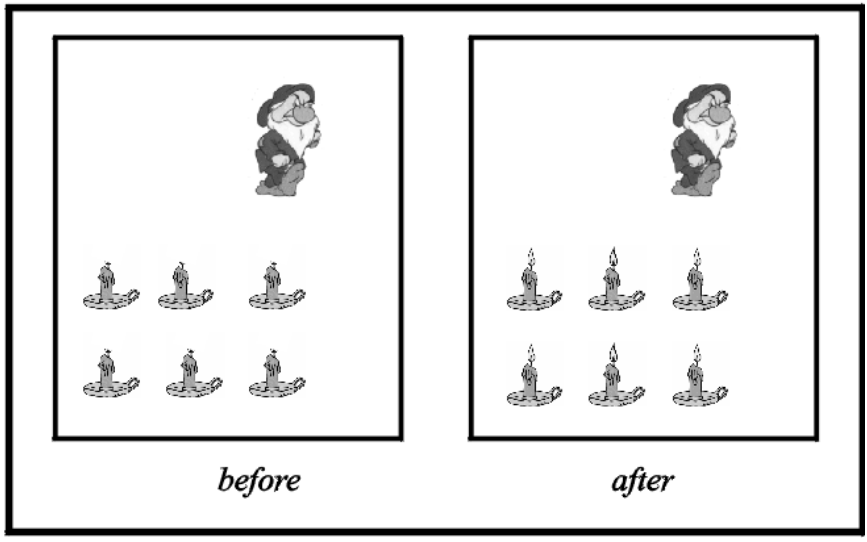

FIGURE 6 Experiment 3: A sample trial (6/6).

empty). Whenever an object was transformed, the change was always made very obvious by color, shape, and so forth. For a full list of materials, see Appendix C.

In the control questions, the sentence quantifier was *all*, and a ratio of 6/6 was presented in the "after" picture. Thus, all control questions were expected to receive *yes* answers.

We assigned each participant randomly to one of two conditions (*most* vs. *half*) in a between-subjects design. In the *most* condition, children's ages ranged from 3;5 to 5;3 in the younger group (M = 4;5), from 5;11 to 8;3 in the middle group (M = 7;2), and from 8;11 to 11;11 in the older group (M = 10;0). In the *half* condition, children's ages ranged from 3;7 to 5;6 in the younger group (M = 4;6), from 6;8 to 8;3 in the middle group (M = 7;5), and from 8;5 to 11;10 in the older group (M = 10;2). Within each condition, we randomly assigned each participant to one of two counterbalanced stimuli presentation orders.

6.3. Results

Results from this experiment are presented in Table 2. (Performance on control items was error free and is not included.) Beginning with *most*, we conducted a repeated measures ANOVA using the number of *yes* responses as a dependent variable with age (adults, younger children, middle children, older children) as a between-subject factor and trial type (0/6, 1/6, 2/6, 3/6, 4/6, 5/6, 6/6) as a within-subjects factor. The analysis yielded a main effect of age, $F(3, 27) = 0.67$, $p = .0005$; a main effect of trial type, $F(6, 27) = 10.43$, $p < .001$; and an interaction between age and trial type, $F(18, 27) = 4.6$, $p < .0001$.

We conducted three separate repeated measures ANOVAs next comparing each group of children to adults, with trial type as a within-subjects variable and age as a between-subjects variable. For the youngest group, the analysis revealed a main effect of age, $F(1, 11) = 2.36$, $p < .0001$; a main effect of trial type, $F(6, 11)$

TABLE 2
Experiment 3: Children's and Adults' Interpretation of *Most* and *Half* (Count Contexts)

Age	*0/6*	*1/6*	*2/6*	*3/6*	*4/6*	*5/6*	*6/6*
% acceptance of *most*							
3- to 5-year-olds	23.33	60.00	93.33	90.00	100.00	100.00	90.00
6- to 8-year-olds	3.70	37.03	37.00	74.00	85.00	88.80	100.00
9- to 11-year-olds	3.70	7.40	25.90	88.80	77.70	77.70	37.00
Adults	0.00	0.00	0.00	30.00	100.00	100.00	90.00
% acceptance of *half*							
3- to 5-year-olds	16.67	70.00	76.67	96.67	80.00	80.00	86.67
6- to 8-year-olds	3.70	14.80	33.30	88.80	66.60	62.90	40.70
9- to 11-year-olds	0.00	0.00	0.00	100.00	7.40	3.70	0.00
Adults	0.00	0.00	0.00	100.00	26.67	23.33	26.67

= 10.43, $p < .0001$; and an interaction between age and trial type, $F(6, 11) = 4.6$, $p < .0001$. Adults accepted ratios over 3/6 as confirming *most*, but very young children were much more lenient in their acceptance responses. Pairwise comparisons of the number of *yes* responses on each trial type revealed that the youngest group's performance was significantly different from adults on the 0/6, 1/6, 2/6, and 3/6 trials (at $p < .05$)—that is, those trials for which the correct response was *no*. In the rest of the trials for which the correct response was *yes*, young children's performance did not differ from adults'.

There was no main effect of age in the analyses comparing adults to either the middle or the older group. However, in both cases, there was a main effect of trial type, $F(6, 10) = 35.40$, $p < .0001$ and $F(6, 10) = 14.28$, $p < .0001$, respectively, and an Age × Trial Type interaction, $F(6, 10) = 1.67$, $p = .03$ and $F(6, 10) = 2.4$, $p = .009$, respectively. Pairwise comparisons showed that the middle group differed from adults in acceptance rates only for the 1/6, 2/6, and 3/6 items (all ps $< .05$). Children in the older group were similar to adults with the exception of 3/6 trials (which they accept much more frequently than adults: 88.8% vs. 30%, respectively, $p < .05$) and the 6/6 trials (which they accept much less frequently than adults: 37% vs. 90%, $p < .05$).

We next took a more detailed look at individual responses to study children's most common error patterns in the comprehension of *most*. We determined each participant's acceptance patterns by looking at whether he or she gave an affirmative response to at least two out of the three questions per ratio. Results are summarized in Table 3. None of the young children showed adultlike performance with *most*: 3 children answered *yes* throughout, and the rest displayed various incorrect interpretations for the quantifier ("more than one"—4 children, "some/at least one"—2 children; "at least half"—1 child). Of the 9 children in the middle group, only 1 child performed without errors, accepting every trial over 3/6; 4 children interpreted *most* as "at least half"; 3 children took *most* to mean "some" or "at least one"; and another child interpreted *most* as "all." Of the 9 older children, only 1 gave legitimate ("more than half") responses (excluding 6/6 trials); 3 children took the quantifier to mean "at least half" (2 of them excluded 6/6 trials); 2 children thought that *most* means "exactly half"; the remaining 3 children interpreted *most* as "more than one" (but in one case excluding 6/6). Finally, of the 10 adults, 6 performed without errors, accepting every trial over 3/6; another one did the same but refused to accept 6/6 trials (still a legitimate response); the remaining 3 adults gave "at least half" responses. These patterns confirm the conclusion that the youngest group of children has not acquired the meaning of *most*; even in the other two groups, children have not achieved adultlike performance with the quantifier.

We now turn to *half*. As Table 2 shows, adults overwhelmingly accepted "exact" (3/6) interpretations for *half* but were much less likely to accept "at least" (over 3/6) interpretations. Especially for the crucial 6/6 trials, acceptance rate for *half* was only 26.67%. Children's acceptance rates, like the adults', were highest for 3/3 trials and the ratios closest to 3/3; the middle and especially the older group were very similar to adults. An initial repeated measures ANOVA using the

TABLE 3
Experiment 3: Distribution of Participants Depending on Response Type

Age	*"More Than Half"*	*"At Least Half"*	*"Exactly Half"*	*"Some/At Least One"*	*"More Than One"*	*"All"*	*All-Yes*
Response patterns for *most*							
3- to 5-year-olds[a]	—	1	—	2	4	—	3
6- to 8-year-olds[b]	1	4	—	3	—	1	—
9- to 11-year-olds[b]	1	3	2	—	3 (1)	—	—
Adults[a]	7 (1)	3	—	—	—	—	—
	"Exactly Half"	*"At Least Half"*	*"Some/At Least One"*	*"More Than One"*	*"Many"*	*All-Yes*	*Other*
Response patterns for *half*							
3- to 5-year-olds[a]	1	1	6	1	—	1	1
6- to 8-year-olds[b]	3	2 (1)	1 (1)	2	1	—	—
9- to 11-year-olds[b]	9	—	—	—	—	—	—
Adults[a]	7	2 (1)	—	—	—	—	1

Note. Number in parentheses shows number of children giving upper bounded responses (preceded by total number of children per response type).
[a]$n = 10$. [b]$n = 9$.

number of *yes* responses as a dependent variable with age as a between-subjects factor and trial type as a within-subjects factor yielded a main effect of age, $F(3, 27) = 1.25$, $p < .0001$; a main effect of trial type, $F(6, 27) = 55.63$, $p < .0001$; and an interaction between age and trial type, $F(18, 27) = 3.61$, $p < .0001$.

We next conducted three separate repeated measures ANOVAs comparing the number of *yes* responses offered by each group of children to those offered by adults using trial type as a within-subjects variable and age as a between-subject variable. As was the case for *most*, for the younger group of children, there was a main effect of age, $F(1, 11) = 1.25$, $p = .0002$; a main effect of trial type, $F(6, 11) = 46.83$, $p < .0001$; and an Age × Trial Type interaction, $F(6, 11) = 3.09$, $p = .0021$. Children in the young group were much more likely than adults to accept all ratios as instances of *half*: Pairwise comparisons reveal that their acceptance rates differed significantly from adults' on all trials ($p < .05$) except for the 3/6 trial.

The analysis comparing adults and children of the middle age group returned a main effect of trial type, $F(6, 10) = 61.76$, $p < .0001$, but no main effect of age or interactions. Nevertheless, pairwise comparisons revealed that this group gave significantly different responses than adults on certain types of trials (2/6, 4/6, and 5/6; $p < .05$). The ANOVA comparing adults and the older group returned no significant main effects or interactions. Pairwise comparisons of acceptance rates on each individual trial type between the older group and adults showed no significant differences.

We now inspect individual responses to *half* trials more closely (see Table 3). Of the 10 young children, only 1 gave "exactly half" responses, and the majority (6) took *half* to mean "some." The remaining 3 children all gave different kinds of wrong answers (1 accepted both "exactly half" and "all" as confirming *half*, another treated *half* as meaning "more than one," and the third gave affirmative responses to all trial types). Of the 9 middle-group children, 3 adopted "exactly half," and another 2 "at least half" construals (for 1 child, the latter excluded 6/6 trials); another 2 children gave "more than one" responses; the remaining 2 children each adopted a different erroneous interpretation ("some but not all" and "many," which corresponded to the 4/6 and 5/6 trials). All of the older children and 7 of the 10 adults gave "exactly half" responses. Two more adults gave "at least half" responses (but 1 of them rejected 6/6 trials), and another one gave overall "at least half" responses (3/6, 4/6, 6/6) but failed to accept 5/6 outcomes. These patterns jointly show that the middle group to a certain extent and certainly the older group have acquired the semantics of *half*.

Last, we turn to a comparison between the results for *most* and *half*. In light of the theoretical issues of interest, we focus on a comparison between children's and adults' performance in the 6/6 trials. Recall that adults' acceptance rates for *most* and *half* on these trials were 90% and 26.67%, respectively (Table 2): This difference is statistically significant ($p = .0016$). The asymmetry persisted in children as soon as they were able to narrow down the meaning of the quantifiers: The

acceptance rate for *most* versus *half* in the 6/6 trials was 100% versus 40.7% for the middle group (a significant difference: $p = .0007$) and 37% versus 0% for the older group (also significant: $p = .01$).[11]

6.4. Discussion

From the point of view of the neo-Gricean versus bilateralist analyses, data from this experiment support two main conclusions. First, in contrast to the predictions of the bilateral view, adults overwhelmingly accepted that a speaker by uttering a statement of the form *most Fs G* could have meant (among other things) *all Fs G*. In other words, *most Fs G* does not seem to semantically exclude *all Fs G*. In that respect, *most* behaves differently from *half*, which excludes higher scalar values more readily. In overall shape, these results confirm those of Experiment 2.[12]

Second, children who are in the process of acquiring the semantics of the quantifiers were more likely to assign upper bounded interpretations to *half* than to *most*. Leaving aside our youngest group, who had not yet acquired the semantics of the two quantifiers, the two other groups seemed to treat *half* as punctual and *most* as a lower bounded range (see below for details on children's early conjectures). As before, the direction of this asymmetry is predicted by the classical view but is unexpected on the revised view on *most*.

From a developmental perspective, the experiment introduces an array of novel data on the acquisition of quantifier meaning that are interesting in their own right. One of the most striking findings is the fact that very young children made massive errors with both *most* and *half* and accepted as instances of the quantifiers outcomes that lie outside the quantifiers' denotation. As these data show, very young children seemed to know that *most* and *half* do not denote "none" because they mostly rejected use of the quantifiers for the 0/6 trials (i.e., when no transformation of the scene had taken place). Nevertheless, children at this stage entertained various nonadult meanings for *most* (mainly "more than one" but also "some/at least one") and *half* (mainly "some").

Even children in the middle and older groups made numerous errors with quantifier denotation. A frequent error seemed to lie in with the lower boundary of *most*: One third or more of these children took *most* to mean "at least half," that is,

[11]Four middle and four older children offered spontaneous (and appropriate) corrections of some of the statements. *Most* was corrected to *one* (two responses), *two* (three responses), *some* (one response) and *all* (four responses); *half* was corrected to *all* once. Interestingly, a 10-year-old said in response to a 2/6 trial, "No, it needs to be at least half for me (to be *most*)."

[12]In Experiment 2, adults accepted *most* in *all* trials only half of the time, whereas in Experiment 3, they did so in the great majority of cases. Similarly for *half*, Experiment 2 led to more upper bounded responses than Experiment 3. The difference was probably due to the wording of the question: "what the speaker said" encourages participants to consider the whole range of outcomes allowed by the semantics of *most* and *half*, thereby giving rise to more "at least" interpretations than would otherwise be obtained.

they accepted 3/6 outcomes as instances of *most*. It is possible that acceptance of the 3/6 trials was due to semantic overextension. Alternatively, children may have known that this ratio lies outside the denotation of *most* but decided that it is close enough for the purposes of accepting *most*. In any case, a look at children's responses reveals some puzzlement at the 3/6 trials. One of the middle-group children who gave adultlike responses commented that a 3/6 trial was not "really *most*." Two other children, one in the middle and one in the older group, who overall gave "at least half" answers seemed unsure whether *most* applied in the 3/6 trials and gave affirmative responses only when pressed to make a choice. It is worth mentioning that 3 adults also accepted *most* in the 3/6 trials; it is unlikely that, for them, the quantifier denotation was unclear or erroneous.

Overall, as Table 3 shows, performance on *half* was markedly better than on *most*. Especially the 9- to 11-year-olds seem to have converged on the adult interpretation of *half* but still made errors with the meaning of *most*. The developmental lag between the two quantifiers makes sense. To compute the truth value of an utterance such as *I'm going to light most/half of the candles*, the hearer needs to calculate the proportion of candles that are lit and compare it to the proportion that are not lit. In the case of *most*, the first set needs to be larger than the second; in the case of *half*, the two sets need to be equal. The computations underlying the comprehenson of *most* thus presuppose (and hence are more complex than) those underlying *half*. Interestingly, as we saw, children accepted *half* when *most* had been uttered and occasionally (in the older group) even misinterpreted *most* to mean "exactly half." We return to the developmental significance of these results in the General Discussion.

7. GENERAL DISCUSSION

7.1. Scalars and the Semantics–Pragmatics Interface

According to the traditional neo-Gricean view, the proportional quantifier *most* is semantically lower bounded but often gives rise to the meaning "not all" through a scalar implicature (Horn (1972)). More recent proposals, however, have claimed that *most* does not generate a scalar implicature but is semantically both lower and upper bounded (Ariel (2003; 2004; in press)). In this article, we reported a series of experiments that sought to evaluate these two different semantic–pragmatic accounts of *most* using data from adults and children.

One set of experiments focused on a comparison between *most* and *half* because, on the classical view, *half* and other exact determiners (e.g., cardinals) should admit bilateral interpretations more readily than the upward-oriented *most*; however, no such difference should exist if *most* is both lower and upper bounded. We found that, in nonbiasing contexts, adults were more likely to treat *most* as being compatible with *all* than *half.* Moreover, a similar asymmetry resurfaced in children's interpretations of the two determiners as soon as children acquired their

semantics. Further experiments compared adults' comprehension of *most* and *more than half.* Even though there are major similarities in the interpretation of the two expressions, *more than half* was more likely than *most* to be treated as compatible with *all.* At first glance, this result appears to offer support to the revised view, which ascribes an upper boundary to *most* but not to *more than half*; however, we showed that these differences can, in fact, be accommodated within the classical theory by exploring the paradigmatic opposition between *most* and *more than half.* Taken together, our studies offer support for the classical, neo-Gricean position on *most* over recent revisions.

In further support of this conclusion, we reconsider briefly two more general assumptions of the bilateral view. Recall that one of the motivations for the bilateralist claim that *most* is not used to implicate "not all" was that such an inference can be detrimental to the speaker's conversational goals. For instance, if the statement *Most people want peace* is used to implicate that not all people want peace, the statement becomes much less effective as an argument for peace. However, this is not the only explanation of the facts. Speakers may well end up implicating things they would be reluctant to encode explicitly (Horn (in press)); furthermore, the previous statement may still be effective in a context in which it is assumed that nobody or maybe a small minority cares about peace.

Another one of the original arguments supporting the reanalysis of *most* came from Ariel's corpus evidence and questionnaire studies showing that a "not all" implicature was rarely conveyed by naturalistic uses of *most* (even though *most* was taken to be upper bounded). However, the standard neo-Gricean account does not, in fact, require that scalar implicatures should arise in the majority of the uses of quantity expressions (but see Levinson (2000) on "default" inferences). The neo-Gricean claim is simply that scalar inferences arise when the stronger scalar term is relevant, and the speaker is expected to know this (cf. the reasoning schema in section 2.1). In discussing our own and others' results, we have emphasized that the computation of scalar inferences depends on whether a stronger statement was indeed relevant and expected given what the speaker knew.

Our explorations of scalar semantics and pragmatics offer a demonstration of ways in which psycholinguistic data from adults and children can confirm and empirically extend theoretical analyses of linguistic phenomena. From a theoretical point of view, it is particularly interesting that our experimental findings reveal an asymmetry between "exact" scalar values (e.g., numbers, *half*) and vague quantifiers (e.g., *some* or *most*), which has been postulated on independent grounds within linguistic theory (see section 2.3).

From a methodological standpoint, these and other studies we have discussed throughout raise the important issue of how participants interpret tasks involving judgments of natural-language sentences (often out of context)—and how in turn the task results should be interpreted by the theorist interested in linguistic meaning. These data add to a growing line of work showing that speakers' intuitions about scalar interpretations vary depending on the nature of the task (Chierchia,

Crain, Guasti, Gualmini, and Meroni (2001), Musolino (2004), Noveck (2001), Papafragou (2003; in press), Papafragou and Musolino (2003)). Furthermore, they confirm that speakers' intuitions about linguistic meaning are subject to pragmatic intrusions (cf. Gibbs and Moise (1997), Koenig (2005)). As a result, these findings raise problems for the view that lexical meaning should be "transparent," that is, (i) accessible to naïve intuition and (ii) straightforwardly related to usage patterns (Ariel (2004)). Once we recognize that lexical semantics is quite abstract, it becomes clear that people's intuitions about the information encoded by natural-language items should not necessarily be respected (Carston (2002), Koenig (2005)). The challenge for both the theorist and the experimentalist is to find the techniques and theoretical machinery to isolate semantic and pragmatic contributions to meaning.

7.2. Acquiring and Processing Scalars

Beyond their theoretical significance for the semantic treatment of scalars, our findings have interesting implications for how scalar expressions are acquired as a class. One of our most robust findings is the fact that children treat quantifiers such as *most* differently from *half* (and potentially other "exact" scalar values such as the cardinals)—hence, they are sensitive to just the semantic differences one would expect given the characterization of these terms in the adult grammar.

Interestingly, recent developmental research from different directions seems to have converged on the same conclusion. In one experiment (Papafragou and Musolino (2003)), 5-year-old children were shown an acted-out story in which a set of three horses jumped over a fence. When children were presented with the statement *Some of the horses jumped over the fence,* they often accepted it as a possible description of the story; however, they were much less likely to accept the statement *Two of the horses jumped over the fence* under the very same circumstances. The asymmetry reveals that numbers tend to receive upper bounded interpretations more than quantifiers—a fact to be expected if numerals, but not quantifiers, have "exact" (albeit underspecified) lexical semantics. Other experiments have shown that children of this age treat *half* as similar to the cardinals, that is, they take it to exclude higher values more than other proportional scalars do. For instance, when presented with a scene in which a girl completes a puzzle, 5-year-olds were much more likely to resist the statement *The girl did half of the puzzle* than the statement *The girl started doing the puzzle*—even though for adults, both utterances implicate noncompletion (Papafragou (2003; in press)).

The number–quantifier asymmetry can be observed at early stages of word learning. Hurewitz, Papafragou, Gleitman, and Gelman (2006) presented 3-year-old children with a set of pictures and asked them to select which picture went with a sentence they heard. Sentences included either numbers (*two* vs. *three*) or quantifiers (*some* vs. *all*). Children were much more successful in finding the pic-

ture that went with a numeral (e.g., *The alligator ate two of the cookies*) than with a quantifier (e.g., *The alligator ate some of the cookies*), even though target pictures and foils were identical across the number and quantifier trials. Interestingly, children tended to confuse *some* and *all* much more than they confused *two* and *three.*

In another related study, Huang, Snedeker, and Spelke (2004) presented 2- to 3-year-old children with two open boxes containing one and three fish as well as a closed box containing an unknown number of fish. Children who had been previously found to know the meaning of the numeral *two* refused to accept that *two* applied to the open three-fish box. When the experimenter asked them to give her a box with two fish, they preferred instead the closed box (guessing that it had to verify an "exactly two" interpretation). This differed strikingly from children's performance with *some* in which there was no preference for upper bounded ("some-but-not-all") interpretations. The empirical generalization from these studies seems to be that quantifiers and numbers/*half* have different semantic and interpretive properties and hence pose different challenges to learners.[13]

Given this general picture, several specific issues remain open about how *most* (and related quantifiers) are acquired and processed. One of the most striking aspects of our data was the fact that our youngest group of 3- to 5-year-olds had considerable problems with quantifier meaning. A natural question to ask is what exactly is the source of this early nonadult performance. A reasonable hypothesis is that at least part of children's difficulty with *most* lies with the fact that it is a higher order quantifier; that is, it involves tracking and comparing the relative size of sets (Barwise and Cooper (1981), Keenan and Stavi (1986)). In a sentence such as *I'm going to color most of the stars blue,* the hearer needs to compute the proportion of the stars that are colored blue and compare it to the proportion that is not colored blue. The comprehension of higher order quantifiers therefore requires nontrivial computational and memory resources (cf. Drozd (2001)). *Half* also relies on the comparison of set sizes but is simpler in that it requires equal set cardinalities.

As our data show, children gradually progress toward recovering from errors and converging on the correct conjectures for quantifier meaning. This pattern is especially clear in the case of *half,* in which 9- to 11-year-olds seemed to have converged on the adult interpretation. With *most,* errors persisted even in the middle and older groups. Children's interpretations had become sensitive to the fact that the denotation of *most* involves a lower bounded range, as we discussed previously; nevertheless, learners had not yet converged on the correct range for the quantifier and sometimes adopted lower bounded denotations that correspond to

[13]This asymmetry argues against the claim that children may use their knowledge of the semantics of quantifiers to bootstrap into knowledge of number meanings (for different versions of this claim, see Bloom (2000), Carey (2001), Spelke and Tsivkin (2001)). For further discussion of this point, see Hurewitz et al. (2006).

"some" or "at least half." Future investigations need to address when children's performance with *most* becomes less unconstrained. They also need to broaden the range of contexts, presentation conditions, and set sizes used to test children's comprehension of quantification.

An interesting question is whether early performance with quantification is influenced by whether the quantifier composes with a count or mass noun (*most/half of the stars* vs. *most/half of the star*). Count and mass uses rely on computation of numerosity versus size/extent, respectively, and might be approached differently by learners. For instance, it could be that the meaning of *most/half* is more transparent when the quantifiers compose with a count noun because the presence of countable objects might help the computation of set relations. Alternatively, calculation of size/extent might be easier and prior to the computation of set cardinality. In another study (Papafragou and Schwarz (2005)), we replicated our earlier Experiment 2 (in which adults interpreted quantifiers in mass contexts) with 3- to 5-year-olds: That study revealed massive lack of understanding of the quantifiers much like the pattern we reported here with data elicited in count contexts. This preliminary evidence seems to indicate that early difficulties with numerical estimation might be robust across mass and count uses of quantifiers.

Finally, our experiments offer evidence for the interpretive preferences of children and adults with respect to different scalar quantifiers. This raises the issue of whether additional interpretations of these quantifiers are available to children and, if so, under which circumstances. For instance, we saw that children (and adults) are more likely to give an exclusive interpretation to *half* ("half and no more") than to *most* ("most but not all"). However, we know from other work that children can cancel this exclusivity inference in linguistic contexts in which such an inference is not licensed (e.g., downward-entailing contexts): For example, in a competition in which prizes go to "whoever draws half a star," children overwhelmingly gave a prize to both (i) contestants who drew just half a star and (ii) contestants who drew a whole star (Papafragou (2003; 2005); cf. also Musolino (2004) on children's flexible interpretation of cardinals). The question now is whether this flexibility extends to nonpunctual *most* (and related quantifiers).

One way of exploring this issue is to ask whether children are sensitive to scalar implicatures from the use of *most* or *some.* Initial evidence on *some* has indicated that children, unlike adults, do not consistently calculate such inferences (Noveck (2001), Papafragou and Musolino (2003)), and our results on *most* seem to point in the same direction. However, other evidence has suggested that children are capable of computing scalar inferences in naturalistic conversational contexts: For instance, they know that an animal that reports that it has eaten some of the oranges implicates that it has not eaten all of them (Papafragou and Tantalou (2004)). What seems to be difficult for children, then, is the computation of a scalar inference in the absence of clear informativeness/relevance demands constrained by speaker intentions. In this study, the environments presented to children left it open whether a stronger alternative (*all*) was relevant and ex-

pected. Under those circumstances, it is not surprising that the contrast with *all* was not used to restrict how *most* should be interpreted. It would be interesting to test how children (and adults) interpret *most* in environments in which the corresponding statements with *all* are clearly relevant and expected.

Another way of exploring pragmatic flexibility is to look at the lower boundary of *most* and its close relatives. Our experiments showed that adults adjusted their interpretation of *most* according to context-driven expectations about what counts as a significant majority. Similar findings on the importance of expected or normal quantities for numerical estimation have emerged for vague quantifiers such as *a few, some,* or *many* (Moxey and Sanford (1993) and discussion in section 4). Further studies have shown that the interpretation of quantifiers is influenced by factors such as the size of the objects in a set or the overall spatial configuration: German speakers assign a smaller number to the German equivalent of *a few* when they hear *A few people are standing before a hut* than when they hear *A few people are standing in front of a building* (Hörmann (1983); cf. Drozd (2001, 357)). They also take *several parcels* to indicate fewer items when they are told the parcels are on a small table than when they are told they are on a large one (Hörmann (1983)). A full account of the acquisition of scalar quantifiers needs to address how children develop the ability to compute such subtle inferences about communicated quantities and how they generalize them across members of the quantificational class.

ACKNOWLEDGMENTS

Portions of this material were developed as part of the second author's MA thesis under the supervision of the first author. We thank Larry Horn, Mira Ariel, Ken Drozd, and two anonymous reviewers for very useful comments. We also thank Alice Huang, Amy Ritter, Laurie Yarzab, and the Early Learning Center in Newark, Delaware, for their help in data collection. Research for this article was partly supported by a General University Research grant from the University of Delaware to Anna Papafragou.

REFERENCES

Ariel, M. (2003) "Does *Most* Mean 'More Than Half'?," *Berkeley Linguistics Society* 29, 17–30.

Ariel, M. (2004) "*Most*," *Language* 80, 658–706.

Ariel, M. (in press) "A 'Just That' Lexical Meaning for *Most*," in K. von Heusinger and K. Turner, eds., *Where Semantics Meets Pragmatics,* Elsevier, Oxford, England.

Atlas, J. (1984) "Comparative Adjectives and Adverbials of Degree," *Linguistics and Philosophy* 7, 347–377.

Barwise, J. and R. Cooper (1981) "Generalized Quantifiers and Natural Language," *Linguistics and Philosophy* 4, 159–219.

Bloom, P. (2000) *How Children Learn the Meaning of Words,* MIT Press, Cambridge, Massachusetts.

Breheny, R. (2005) "A New Look at the Semantics and Pragmatics of Numeral Expressions," ms., University College London.

Carey, S. (2001) "Whorf Versus Continuity Theorists: Bringing Data to Bear on the Debate," in M. Bowerman and S. Levinson, eds., *Language Acquisition and Conceptual Development,* Cambridge University Press, Cambridge, England.

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

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

Carston, R. (2002) *Thoughts and Utterances: The Pragmatics of Explicit Communication,* Blackwell, Oxford, England.

Chierchia, G. (2004) "Scalar Implicatures, Polarity Phenomena, and the Syntax/Pragmatics Interface," in A. Belleti, ed., *Structures and Beyond,* Oxford University Press, Oxford, England.

Chierchia, G., S. Crain, M.-T. Guasti, A. Gualmini, and L. Meroni (2001) "The Acquisition of Disjunction: Evidence for a Grammatical View of Scalar Implicature," *Proceedings from the Annual Boston University Conference on Language Development* 25, 157–168.

Chomsky, N. (1965) *Aspects of the Theory of Syntax,* MIT Press, Cambridge, Massachusetts.

Crain, S. and C. McKee (1985) "The Acquisition of Structural Restrictions on Anaphora," *Proceedings of NELS* 15, 94–110.

De Morgan, A. (1847) *Formal Logic,* Taylor & Walton, London.

Drozd, K. (2001) "Children's Weak Interpretations of Universally Quantified Questions," in M. Bowerman and S. Levinson, eds., *Language Acquisition and Conceptual Development,* Cambridge University Press, Cambridge, England.

Fox, D. (2003) "Implicature Calculation, *Only,* and Lumping: Another Look at the Puzzle of Disjunction," talk handout, Yale University, New Haven, Connecticut.

Gazdar, G. (1979) *Pragmatics: Implicature, Presupposition, and Logical Form,* Academic Press, New York.

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

Gibbs, R. and J. Moise (1997) "Pragmatics in Understanding What Is Said," *Cognition* 62, 51–74.

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

Hamilton, Sir William, of Edinburgh (1860) *Lectures on Logic,* Blackwood and Sons, Edinburgh, Scotland.

Hirschberg, J. (1985) *A Theory of Scalar Implicature,* Doctoral dissertation, University of Pennsylvania, Philadelphia.

Hörmann, H. (1983) "The Calculating Listener, or: How Many Are *Einige, Mehrere*, and *Ein Paar* (*Some, Several* and *A Few*)?" in R. Bäuerle, C. Schwarze, and A. von Stechow, eds., *Meaning, Use, and Interpretation of Language,* de Gruyter, Berlin, Germany.

Horn, L. (1972) *On the Semantic Properties of the Logical Operators in English,* Doctoral dissertation, University of California, Los Angeles.

Horn, L. (1989) *A Natural History of Negation,* University of Chicago Press, Chicago.

Horn, L. (1992) "The Said and the Unsaid," proceedings from *SALT* II, 163–192, Department of Linguistics, Ohio State University.

Horn, L. (in press) "The Border Wars: A Neo-Gricean Perspective," in K. von Heusinger and K. Turner, eds., *Where Semantics Meets Pragmatics,* Elsevier, Oxford, England.

Huang, Y. T., J. Snedeker, and E. Spelke (2004) "What Exactly Do Numbers Mean?" paper presented at the 26th Annual Meeting of the Cognitive Science Society, Chicago.

Hurewitz, F., A. Papafragou, L. Gleitman, and R. Gelman (2006) "Asymmetries in the Acquisition of Numbers and Quantifiers," *Language Learning and Development* 2, 77–96.

Keenan, E. (1996) "The Semantics of Determiners," in S. Lappin, ed., *The Handbook of Contemporary Semantic Theory,* Blackwell, Oxford, England.

Keenan, E. and J. Stavi (1986) "A Semantic Characterization of Natural Language Determiners," *Linguistics and Philosophy* 9, 253–326.

Koenig, J.-P. (1991) "Scalar Predicates and Negation: Punctual Semantics and Interval Interpretations," *Proceedings of the Parasession on Negation of the 27th meeting of the Chicago Linguistics Society*, 130–144.

Koenig, J.-P. (2005) "Mostly a Field Guide," talk delivered at the 79th Annual LSA Meeting, San Francisco, California.

Kratzer, A. (2003) "Scalar Implicatures: Are There Any?" talk delivered at the Milan Workshop on Scalar Phenomena, Polarity and Implicature, University of Milan-Biccoca, Italy.

Lasersohn, P. (1999) "Pragmatic Halos," *Language* 75, 522–551.

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

McCawley, J. (1981) *Everything That Linguists Have Always Wanted to Know About Logic But Were Ashamed to Ask,* University of Chicago Press, Chicago, Illinois.

Moxey, L. and A. Sanford (1993) *Communicating Quantities: A Psychological Perspective,* Lawrence Erlbaum Associates, Inc., Hillsdale, New Jersey.

Musolino, J. (2004) "The Semantics and Acquisition of Number Words: Integrating Linguistic and Developmental Perspectives," *Cognition* 93, 1–41.

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

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

Papafragou, A. (in press) "From Scalar Semantics to Implicature: Children's Interpretation of Aspectuals," *Journal of Child Language.*

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

Papafragou, A. and N. Schwarz (2005) "*Most* Wanted: Explorations in Scalar Semantics," paper presented at the 79th Annual LSA Meeting, San Francisco, California.

Papafragou, A. and N. Tantalou (2004) "Children's Computation of Implicatures," *Language Acquisition* 12, 71–82.

Peterson, P. (1979) "On the Logic of 'Few,' 'Many,' and 'Most'," *Notre Dame Journal of Formal Logic* 20, 155–179.

Pinker, S. (1989) *Learnability and Cognition: The Acquisition of Argument Structure,* MIT Press, Cambridge, Massachusetts.

Sadock, J. (1984) "Whither Radical Pragmatics?" in D. Shiffrin, ed., *Georgetown University Roundtable on Language and Linguistics,* Georgetown University Press, Washington, DC.

Sauerland, U. (2004) "Scalar Implicatures in Complex Sentences," *Linguistics and Philosophy* 27, 367–391.

Spelke, E. and S. Tsivkin (2001) "Initial Knowledge and Conceptual Change: Space and Number," in M. Bowerman and S. Levinson, eds., *Language Acquisition and Conceptual Development,* Cambridge University Press, Cambridge, England.

Sperber, D. and D. Wilson (1985–1986) "Loose Talk," *Proceedings of the Aristotelian Society* 86, 153–171.

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

Submitted 20 February 2004
Final version accepted 7 September 2005

APPENDIX A

Materials for Experiment 1

The sentences in italics were included in the biasing context condition but not in the neutral condition.

All (control)

Here's Happy. He said, "*I'm the best at drawing. Nobody can do it better than me.* I'm going to color all of the star blue."
Here's Grumpy. He said, "*I'm the best at lighting candles. Nobody can do it better than me.* I'm going to light all of the candles."
Here's Snow White. She said, "*I love red fingernails.* I'm going to paint all of my fingernails red."
Here's Sleepy. He said, "*I'm the best at cleaning cars. Nobody cleans cars as well as me.* I'm going to clean all of the cars."
Here's Dopey. He said, "*I'm the best helper.* I'm going to fill all of this pitcher with milk."
Here's Sneezy. He said, "*I'm a great painter. I paint better than anybody.* I'm going to paint all of the wheel brown."

0%

Here's Sneezy. He said, "*I love Snow White. I'm going to give her a surprise.* I'm going to fill most of her closet with clothes."
Here's Snow White. She said, "*I need to wash a lot of dishes with soap.* I'm going to refill most of this bottle with soap."
Here's Bashful. He said, "*I love to drink tea.* I'm going to fill most of this mug with tea."

25%

Here's Sleepy. He said, "*I'm the best at coloring. Nobody colors as well as me.* I'm going to color most of the flag red."
Here's Dopey. He said, "*I love collecting pebbles. It's my favorite hobby.* I'm going to fill most of this aquarium with pebbles."
Here's Doc. He said, "*I'm the best babysitter. I take care of babies so well.* I'm going to fill most of this bottle with milk."

40%

Here's Happy. He said, "*I love hiking. I'm the best at filling the water bottle for my hikes.* I'm going to fill most of this bottle with water."

Here's Sneezy. He said, "*I'm the best at yard work. Nobody does it as well as me.* I'm going to plant most of the hill with grass."
Here's Sneezy. He said, "*I'm the best at making cakes.* I am going to decorate most of the cake for Snow White."

50%

Here's Bashful. He said, "*I really love my neighbors.* I'm going to cut most of this pie and give it to the neighbors."
Here's Bashful. He said, "*I'm really hungry. I'm going to make 10 loaves of bread.* I'm going to use most of the flour to bake bread."
Here's Doc. He said, "*I really love gumballs.* I'm going to fill most of the gumball machine with gumballs."

60%

Here's Dopey. He said, "*I love writing long stories. I write 20-page stories all the time.* I'm going to fill most of this page with writing."
Here's Grumpy. He said, "*I'm the best at making pizza, and I love toppings.* I'm going to put toppings on most of the pizza."
Here's Sneezy. He said, "*I'm really thirsty, and I'm the best at pouring. Nobody pours as well as me.* I'm going to fill most of this glass with juice."

75%

Here's Happy. He said, "*My fish needs a lot of water.* I'm going to fill most of this fishbowl with water."
Here's Dopey. He said, "*I love flower petals. They smell great.* I'm going to pluck most of the petals."
Here's Happy. He said, "*I'm the best gardener.* I'm going to plant most of the garden."

APPENDIX B

Materials for Experiment 2

Most condition (for the other two conditions, *most* was replaced by *half/more than half*)

All (control)

Here's Doc. He said, "I'm going to untie the bow."
Here's Bashful. He said, "I'm going to put another block on top."

Here's Sleepy. He said, "I'm going to cut the apple."
Here's Dopey. He said, "I'm going to draw a heart."

0%

Here's Sneezy. He said, "I'm going to give Snow White a surprise. I'm going to fill most of her closet with clothes."
Here's Bashful. He said, "I'm going to fill most of this mug with tea."
Here's Grumpy. He said, "I'm going to eat most of this cookie right now."

25%

Here's Sleepy. He said, "I'm going to color most of the flag red."
Here's Dopey. He said, "I'm going to collect pebbles. I'm going to fill most of the aquarium with pebbles."
Here's Doc. He said, "I'm going to fill most of this bottle with milk."

40%

Here's Happy. He said, "I'm going to fill most of the water bottle for my hike."
Here's Sneezy. He said, "I'm going to plant most of the hill with grass."
Here's Sneezy. He said, "I am going to decorate most of the cake for Snow White."

50%

Here's Bashful. He said, "I'm going to cut most of this pie and give it to the neighbors."
Here's Bashful. He said, "I'm going to use most of the flour to bake bread."
Here's Doc. He said, "I'm going to fill most of the gumball machine with gumballs."

60%

Here's Dopey. He said, "I'm going to fill most of this page with writing."
Here's Grumpy. He said, "I'm going to put toppings on most of the pizza."
Here's Sneezy. He said, "I'm going to fill most of this glass with juice."

75%

Here's Happy. He said, "I'm going to fill most of this fishbowl with water."
Here's Dopey. He said, "I'm going to pluck most of the petals."
Here's Happy. He said, "I'm going to plant most of the garden."

100%

Here's Happy. He said, "I'm going to color most of the star blue."
Here's Dopey. He said, "I'm going to fill most of the pitcher with milk."
Here's Sneezy. He said, "I'm going to paint most of the wheel brown."

APPENDIX C

Materials for Experiment 3

All (control)

Here's Happy. He said, "I'm going to color all of the stars blue."
Here's Grumpy. He said, "I'm going to eat all of these cookies right now."
Here's Dopey. He said, "I'm going to give all of these kittens a kiss."

0/6

Here's Sneezy. He said, "I'm going to give Snow White a surprise. I'm going to put away half/most of her clothes into the closet."
Here's Snow White. She said, "I'm going to refill half/most of these bottles with soap."
Here's Bashful. He said, "I'm going to fill up half/most of these mugs with hot chocolate."

1/6

Here's Doc. He said, "I'm going to put hats on half/most of these puppies."
Here's Sleepy. He said, "I'm going to color half/most of these flags red."
Here's Dopey. He said, "I'm going to collect half/most of these rocks and put them in the box."

2/6

Here's Happy. He said, "I'm going to fill up half/most of these water bottles for my camping trip."
Here's Sneezy. He said, "I'm going to put half/most of these birds in cages."
Here's Sneezy. He said, "I'm going to put icing on half/most of these cupcakes for Snow White."

3/6

Here's Bashful. He said, "I'm going to take away half/most of these pies and give them to the neighbors."

Here's Bashful. He said, "I'm going to use half/most of these bags of flour to bake bread."
Here's Doc. He said, "I'm going to blow up half/most of these balloons."

4/6

Here's Dopey. He said, "I'm going to draw on half/most of these pages."
Here's Grumpy. He said, "I'm going to eat half/most of these slices of pizza."
Here's Sneezy. He said, "I'm going to open half/most of these presents."

5/6

Here's Happy. He said, "I'm going to fill up half/most of these fishbowls."
Here's Dopey. He said, "I'm going to decorate half/most of these Christmas trees."
Here's Happy. He said, "I'm going to put half/most of these flowers in pots."

6/6

Here's Grumpy. He said, "I'm going to light half/most of these candles."
Here's Sleepy. He said, "I'm going to clean half/most of these cars."
Here's Sneezy. He said, "I'm going to paint half/most of these wheels brown."