Meaning and Context in Children's Understanding of Gradable Adjectives

Kristen Syrett[a], Christopher Kennedy[b], Jeffrey Lidz[c]

[a] Rutgers University, [b] University of Chicago, [c] University of Maryland

Abstract

This work provides the first experimental evidence that adults and children as young as three years of age distinguish between three types of Gradable Adjective (GA) meanings (relative, absolute maximum standard, and absolute minimum standard) based on the role of the context in setting the standard of comparison. While relative GAs such as *big* depend on the context for the standard of comparison, absolute GAs (e.g., *full*, *straight*; *spotted*, *bumpy*) do not. Evidence comes from a pragmatically-oriented task in which we exploit participants' awareness of the existence and uniqueness presuppositions associated with singular definite descriptions and measure their willingness to accommodate when these presuppositions are violated. An analysis of children's reaction times when they eventually accept a puppet's infelicitous request involving a maximal standard GA (*full*, *straight*) suggests a distinction between a pragmatic principle of imprecision and a semantic conception of vagueness.

Meaning and Context in Children's Understanding of Gradable Adjectives

1. Introduction

In this paper, we explore what children and adults know about three specific ways that meaning and context interact: the interpretation of expressions whose meanings are inherently linked to the context (semantic context dependence); conditions on the felicitous use of expressions in a discourse context (presupposition accommodation); and informative uses of expressions in contexts in which they strictly speaking do not apply (imprecision). Our empirical focus is on the appearance of gradable adjectives in singular definite descriptions, the use of which is only felicitous as long as the pragmatic presuppositions introduced by the definite determiner are satisfied. We show that by three years of age, children are sensitive to all three varieties of context/meaning interaction, and that their knowledge of this relation with the definite description is appropriately guided by the semantic representations of the gradable adjective appearing in it. These findings thus suggest that children's semantic representations of the gradable adjectives we investigated and the definite determiner *the* are adult-like and that they are aware of the consequences of these representations when relating meaning and context. Bolstered by adult participant responses, this work provides important experimental support for theoretical claims regarding the semantics of gradable predicates and the nature of different types of "interpretive uncertainty", specifically vagueness vs. imprecision.

2 Background

2.1 Relative Gradable Adjectives, Context Dependence, and Accommodation

Gradable adjectives (GAs) are adjectives whose core meanings involve a relation to a scalar concept on the basis of which objects can be ordered (e.g., height, weight, cost, etc.). Distributionally, GAs are identified by the fact that they can appear felicitously in comparative constructions (e.g., *taller than*, *as heavy as*, *less expensive than*) and with

various types of degree morphemes (measure phrases, intensifiers, etc.) whose function is to specify where the argument of a GA is located on the corresponding scale. Our focus in this paper is the meaning and use of GAs in the positive form, which lack any overt degree morphology: examples such as *(is) tall, (seems) heavy* and *(an) expensive (book),* and so forth. From a semantic perspective, the positive form is interesting because its meaning is (typically) tied to the context: what counts as tall can vary, so the extension of *tall* is correspondingly context dependent. Contextual factors relevant to the calculation of the extension of the positive form include (but are not limited to): the denotation of a modified noun (*tall snowman/building/mountain/*etc.), an explicit or implicit comparison class (*tall for a gymnast*), extralinguistic knowledge (e.g., that a snowman described as 'tall' was built by third graders vs. fraternity brothers; see Kamp and Partee (1995)), and the interests and expectations of the participants in the discourse (Fara 2000).

In what follows, we will refer to GAs that have context-dependent positive forms as *relative GAs* (to be distinguished later in section 2.2 from a second class of GAs that do not). A common analysis of the positive form of relative GAs is that they denote properties that are true of an object just in case it can be related to a degree of the scalar concept encoded by the GA that exceeds a contextually determined STANDARD OF COMPARISON (see e.g., Bartsch and Vennemann, 1972; Cresswell, 1976; Kennedy, 1999, 2007; Klein, 1980, 1991; Ludlow, 1989; von Stechow, 1984; Wheeler, 1972; *inter alia*). The standard of comparison is a degree that typically corresponds to an 'average' or 'norm' for the scalar concept relative to some salient set (a comparison class), which, as mentioned above, may be explicitly indicated, may be inferred based on other information in the sentence, or may be implicit.

There are various compositional implementations of this kind of analysis, which differ primarily in the semantic type assigned to gradable adjectives and in more general assumptions about the relation between the morphologically unmarked positive form of a GA

and the various forms with explicit degree morphology. For the purposes of this paper, we will adopt a variant of the analysis in Kennedy (2007), in which the core denotation of a gradable adjective is a function from objects to degrees (a measure function), and the positive form is the result of combination with a null degree morpheme *pos* that has the denotation in (1) (see also Bartsch and Vennemann, 1972).[1]

(1) $[\![\textit{pos}]\!]^c = \lambda g_{\langle e,d \rangle} \lambda x_e . g(x) \geq \mathbf{s}(g)(c)$

Here **s** is a context sensitive function that returns a degree on the scale used by g (its range) in context c that represents an appropriate standard of comparison for the kind of measurement that g expresses. Thus if the adjective *tall* denotes the measure function **tall** (a function from objects to positive degrees of height), the denotation of the positive form of the adjective is the context dependent property in (2), which is true of an object in context c iff its height exceeds the standard of comparison for height in c.

(2) $[\![\textit{pos}\ \text{tall}]\!]^c = \lambda x . \mathbf{tall}(x) \geq \mathbf{s}(\mathbf{tall})(c)$

To see how this analysis allows variation in the extension of the positive form, let us consider the evaluation of a sentence like *Anna is tall* in two different contexts: *c(gymnast*), in which we are talking about (female) gymnasts, and *c(woman)*, in which we are talking about women in general. If the standards of comparison in these two contexts are distinct – and in particular if the standard in *c(gymnast)* is lower than the standard in *c(woman)*, as will normally be the case – then given the semantics of *pos*, the set of things that the predicate is true of in the former context is a subset of the set of things it applies to in the latter context. If Anna's height falls in between the standard of comparison in *c(gymnast)* and *c(woman*), as represented graphically in (3), then given (1), Anna is in the extension of $[\![\textit{pos}\ \text{tall}]\!]^{c(gymnast)}$ but is not in the extension of $[\![\textit{pos}\ \text{tall}]\!]^{c(woman)}$, and the truth of (2) varies accordingly.

[1] Degrees are elements of scales, which are triples <D, <, δ> such that D is a set of degrees, < is a total ordering on D, and δ is a dimension (e.g., HEIGHT, TEMPERATURE, etc.).

(3) HEIGHT: 0 ---------- **s**(**tall**)(*c(gymnast)*) ----- **tall**(**anna**) ----- **s**(**tall**)(*c(woman)*) ------->

In what follows, we will generally use the adjective alone to refer to the positive form; that is, we will say *tall* rather than *pos tall* for perspicuity, unless it is important to distinguish the compositional details of the positive form from some other degree-modified form.

The context dependence of relative GAs means that there is a great deal of flexibility in their use: the same GA can have different extensions in different contexts, while expressing the same property. A particularly striking example of this flexibility comes from the use of relative GAs in definite descriptions, which is the construction we will focus on in this paper.[2] Definite descriptions have been the focus of a great deal of work in semantics, pragmatics, and the philosophy of language, most of which converges in some form or another on two central semantic/pragmatic claims: a singular definite DP of the form *the* ϕ carries with it two presuppositions. First, it presupposes that there is an object that satisfies the property encoded by ϕ (the EXISTENCE presupposition). Second, it presupposes that the object uniquely satisfies ϕ (the UNIQUENESS presupposition). (See Abbott, 1999; Birner and Ward, 1994; Heim, 1990; Kadmon, 1990; Neale, 1990; Roberts, 2003; Russell, 1905; Strawson,1950; and many others for representative discussion.) There are important differences of opinion on the relation between these conditions, the extent to which they can be overridden, and the way that apparent violations should be handled, but for our purposes we can take them as reasonable approximations of what a speaker is committed to in order to felicitously use a definite description, and what a hearer takes to be the case when accepting its use.

The effect of these presuppositions on judgments of felicity can be illustrated by the following example. Consider a situation in which two individuals A and B are sitting across

[2] See Kyburg and Morreau (2000) for additional discussion of the semantics of such constructions and their implications for dynamic semantics.

from each other at a table, there are two blue rods of unequal lengths on the table in front of B, and A's goal is to get B to pass over one of the rods. In such a context, A cannot felicitously use (4a) to make this request, because the uniqueness presupposition of the definite description *the blue rod* is not met: there are two objects in the context that satisfy the property *blue rod*. (We use the '#' symbol to indicate infelicity.) By the same token, A's utterance of (4b) would also be infelicitous, in this case because the existence presupposition is not met: there is no object that satisfies the property *red rod* in the context. Speaker A *can*, however, felicitously use (4c) to request the longer of the two rods.

(4) a. #Please give me the blue rod.

b. #Please give me the red rod.

c. Please give me the long rod.

Importantly, (4c) is felicitous regardless of whether the two rods are independently judged to be both long, both not long, or one long and the other not: all that matters is that there is a difference in length between them.[3]

[3] This is a bit of an oversimplification. As pointed out in Kennedy (2007, p. 19), the difference in length must be greater than a certain threshold: *the long rod* is infelicitous as a description of the longer of two rods that differ by only a small (but noticeable) degree, in contrast to an explicit comparative like *the longer rod*. This fact is presumably due to a second feature of relative GAs in the positive form, which we are setting to the side here: vagueness. Vagueness manifests itself both as an inability or unwillingness on the part of speakers to judge some objects in the domain of the predicate as (not) having the property (so-called 'borderline cases'), and as an inability or unwillingness to distinguish between objects that are very similar to each other relative to the scalar property that the predicate encodes. The latter difficulty is what underlies the 'threshold effect' in definite descriptions, as well as the more commonly discussed judgments about the inductive premise of the Sorites

The crucial difference between (4c) and (4a-b) is that *long* is a positive form relative GA, and so denotes the context-dependent property in (5), which is true of an object just in case its length exceeds the contextual standard for length in the context of utterance.

(5) $\lambda x.\mathbf{long}(x) \geq \mathbf{s}(\mathbf{long})(c)$

In the situation described above, the presuppositions introduced by A's use of the definite description *the long rod* require there to be a unique long rod in the context. Given that there are two salient rods in the context, the only way to satisfy these presuppositions is to make *long* true of one of them and false of the other one. Because the rods have unequal lengths, this result can be achieved by positing a standard that differentiates between the rods, as represented in (6).

(6) LENGTH: 0 ---------- **long**(**rod1**) ----- **s**(**long**)(*c*) ----- **long**(**rod2**) ---------->

Crucially, such a standard is posited automatically, as part of the discourse interpretation of the definite, regardless of whether the rods are independently judged to be long or not – i.e., independently of the prevailing standard of length (for rods). This is thus an instance of presupposition accommodation (Lewis, 1968, 1971), whereby a new standard is accommodated in order to make the discourse model consistent with the presuppositions of the definite description.[4]

Paradox (e.g., "any rod that is just a tiny bit shorter than a long rod is also long"). A full account of the semantics of the positive form needs to take both vagueness and context dependence into account (and indeed many accounts of the former make crucial use of the latter; see e.g., Bosch, 1983; Fara, 2000; Kamp, 1981; Kamp and Partee, 1995; Raffman, 1994, 1996; Soames, 1999), but since our focus here is on relatively clear cases of context-dependent shifts in extension, we will set aside questions about vagueness.

[4] While accommodation of a standard in definite descriptions is in general quite flexible, some adjectives do impose additional constraints. For example, more evaluative adjectives

2.2 Absolute Gradable Adjectives and Imprecision

As it turns out, not all GAs have context-dependent denotations in the positive form. In addition to the large set of relative GAs, there is a class of adjectives that are demonstrably gradable but which have been argued to have fixed, context independent standards. Following Unger (1975), we refer to this class as *absolute* gradable adjectives (see also Rusiecki, 1985; Rotstein and Winter, 2004; Kennedy and McNally, 2005, Kennedy, 2007). Previous researchers have argued for a distinction between absolute and relative GAs on the basis of modifier selection and entailment patterns (see Rusiecki, 1985; Cruse, 1986; Rotstein and Winter, 2004; Kennedy and McNally, 2005; Kennedy, 2007; and Syrett, 2007), and have identified two distinct classes. *Maximum standard absolute GAs* such as *full, straight, flat, dry,* and *pure* require their arguments to possess a maximal degree of the relevant property (*full* is true of a container just in case it is completely full; *straight* is true of an object just in case it has no bend). In contrast, m*inimum standard absolute GAs* such as *spotted, bumpy, wet, bent* and *impure* merely require their arguments to merely possess some nonzero degree of a gradable property (*spotted* is true of an object as long as it has some spots; *bent* is true of an object if it has some degree of bend).

Evidence that absolute GAs are gradable comes from the fact that they combine with comparative and other degree morphology: we can talk about a container that is fuller than another, a line that is not straight enough, a tie that is less spotted than a shirt, or a pole that is too bent for a tent. Such facts indicate that absolute GAs have the same core semantic type as

---

like *fat* or *pretty* and polar-negative adjectives like *short* tend to impose a general requirement that their arguments have the property they measure above a certain threshold. Such markedness effects hold across constructions, however, showing up in morphological comparatives and other degree constructions, as well as the kinds of constructions we are focusing on here.

relative GAs; on our analysis, this means that they denote measure functions. We can further account for the non-context-dependent interpretations of the positive form by positing denotations like (7a-b) for maximum and minimum standard GAs, respectively (here exemplified by the denotations for positive form *full* and *spotted*), where **max** and **min** are functions that return the maximum and minimum degrees on the scales that constitute the ranges of their measure function arguments.

(7) a. $\lambda x.$**full**(x) = **max**(**full**) (maximum standard GA)

b. $\lambda x.$**spotted**(x) > **min**(**spotted**) (minimum standard GA)

Note that these denotations are not what we would expect if the positive form of an absolute GA is derived in the same way as that of a relative GA, through combination with *pos*. Instead, we should expect a context dependent denotation of the sort we posited for *tall* and *long* in (2) and (5). Kennedy (2007) has argued that truth conditions equivalent to (7a-b) can be derived for absolute adjectives using the denotation of *pos* given above in (1), by fleshing out the details of the standard-identifying function **s** in a way that takes into account the scalar properties of the adjective. However, since our focus here is not on how the truth conditions represented in (7a-b) are actually derived, but rather on the predictions about the meaning and acceptability of certain uses of absolute GAs given such truth conditions, we will make the simplifying assumption that the positive forms of relative GAs and maximum/minimum standard GAs are distinct in the way presented here (e.g., because of an ambiguity in *pos*).[5]

[5] Note, however, that we maintain the core assumption that all gradable adjectives have the same basic type: they denote measure functions of type <e,d> (or alternatively, degree relations of type <d,et>). As such, they may combine with the full range of degree morphology, modulo any constraints based on scalar properties (see Rotstein and Winter, 2004 and Kennedy and McNally, 2005). For example, composition of *full* with a comparative

Specifically, we are interested in two ways that absolute GAs can help us understand how much children know about context-dependent aspects of meaning. The first involves their use as controls for an investigation of the definite description data discussed in the previous section, involving presupposition accommodation via contextual shifting of a standard of comparison. Suppose we discover that children correctly associate e.g. *the long rod* with the longer of two rods across a range of contexts and rod lengths. We cannot actually conclude from this fact alone that they have acquired a context-dependent, relative meaning for *long*, because it is possible, especially for very young children (cf. Nelson and Benedict, 1974), that they are misanalyzing morphologically positive *long* as semantically comparative *longer*. If this were the case, we would not be able to conclude anything about their knowledge of the positive form.

Absolute GAs provide a means of testing for this possibility. If children analyze the adjective in a definite description of the form *the A NP* as semantically positive, then relative and absolute GAs should pattern differently in contexts in which *the A NP* is used to distinguish between two objects that manifest different degrees of *A*. For a relative GA, children should always accept the description as applied the higher-ranked object, as outlined above. However, for a maximum standard absolute GA, children should reject the description in contexts in which both objects fall below the maximum standard (e.g., *the full jar* used to describe the fuller of two observably non-full jars), as this violates the existence presupposition of the definite given (7a); and for a minimum standard absolute GA, children should reject the description in contexts in which both objects fall above the minimum standard (e.g., *the spotted disk* used to describe the more spotted of two observably spotted

---

[*er than this cup*] results in a property that is true of an object just in case its fullness (a measure of the amount of stuff it contains) is greater than that of the cup, which may be true even if it is not full in the absolute sense.

disks), as this violates uniqueness given (7b).[6] In contrast, if children analyze the adjective as semantically comparative, relative and absolute GAs should pattern the same in these contexts, since *the Aer NP* can always be used to identify the higher-ranked object, regardless of its relation to the standard associated with the corresponding positive form.

The second reason that we are interested in absolute GAs is because they may provide insights into subtly different ways that context interacts with meaning and use. Our starting point is intuitions about the meanings of maximum standard absolute GAs like *full*. Even though the literature cited above has provided a number of arguments in favor of the maximum standard denotation in (7a), there is nevertheless an initial intuition that this meaning is too strong, and that *full* instead merely requires that an object be 'close to full', allowing for different approximations to suffice in different contexts. For example, while it is clear that a jar that is only half full cannot be truthfully described as *a full jar*, and maybe not one that is 3/4 full either, judgments can become murkier when when we consider a jar that is 7/8 or 15/16 full, or any amount that is extremely close to being ful. At some point we would typically be willing to start calling the jar *full* (and would therefore be willing to refer to it as *the full jar*), and this point might be different in different contexts (based on our goals, the types of jars, the contents of the jars, etc.).

One interpretation of this fact is that the denotation in (7b) is incorrect, and instead *full* should be assigned a context-dependent meaning just like *long* or *tall* (see Lewis 1979). However, another possibility is that (7b) is correct, and this fact indicates that speakers are willing to tolerate imprecision: use of a sentence or description that is false but 'close enough to true' for the purposes of the conversational exchange (Lasersohn, 1999). The experiments

[6] This is the predicted pattern of behavior for adults. Since our study uses adult controls, a secondary result will be to provide experimental support for the claim that absolute GAs do in fact have the denotations in (7a-b).

we report in this paper provide new data relevant to these issues, because they indicate that (for children, at least) uses of relative GAs in which a standard of comparison is shifted to accommodate the presuppositions of a definite DP are processed differently from apparently similar, 'imprecise' uses of absolute GAs. This suggests that apparent context sensitivity manifested by the two classes of GAs involves different kinds of context/meaning relations: relative GAs are context-sensitive relative to an aspect of their *meaning* (variability in the standard of comparison), while absolute GAs are context-sensitive relative to their *use* (tolerance of imprecision).[7]

2.3 Previous Work on Children's Understanding of Gradable Adjectives and Definite Descriptions

Previous lines of research have focused on relative GAs such as *big*, *tall*, *high*, and *low*, demonstrating that preschoolers allow the standard of comparison to shift. A series of papers by Gelman and Ebeling (Ebeling and Gelman, 1988, 1994; Gelman and Ebeling, 1989) showed that that children's ability to judge the size of an object and their facility to move between standards is affected by the nature of the standard of comparison. Smith and her colleagues found that children make use of the given range of values when assigning the standard of comparison (Smith, Cooney, and McCord, 1986) and identified some of the factors affecting children's willingness to relabel an object once the standard has shifted (Sera and Smith, 1987). Finally, Barner and Snedeker (2007, 2008) showed that children at age four shift the standard towards the appropriate pole when additional same-kind objects of different sizes are included in a relevant set. Together, these studies demonstrate that

[7] A question that we will not address in this paper concerns the factors that determine whether a GA is relative or absolute, as these are discussed in detail in other work (see Rusiecki, 1985; Rotstein and Winter, 2004; Kennedy and McNally, 2005; Kennedy, 2007; Syrett, 2007).

preschoolers take a range of contextual information into account when assigning the standard of comparison for these adjectives. However, as none of these studies included non-relative GAs as controls (see section 1.3 above), it has remained an open question whether children consistently take contextual information into account and shift the standard of comparison for *all* adjectives (gradable or not), or appropriately restrict this behavior to relative GAs.

The study that has come closest to introducing a control set of adjectives is Nelson and Benedict (1974), in which children were asked for their judgments about properties of objects instantiated in a series of pictures. When asked about the comparative form of certain adjectives (e.g., *taller*), children exhibited a reaction time more similar to that observed with the positive form of that same adjective than they did with the comparative forms of other adjectives (e.g., *happier* or *leafier*). The authors interpreted these results as demonstrating that only the positive form of the former adjectives carry with them an implicit comparison, and thus it takes no more time to process the comparative form. While these results are suggestive, because the authors imposed distinctions among different classes of adjectives that do not directly correspond to semantic distinctions found in natural language, it is difficult to draw conclusions from their results. Our study aims in part to fill this gap in literature. By starting with the theoretically-motivated semantic distinction between relative and absolute GAs discussed in sections 1.2-1.3, we are able to bolster our claims about what children know about relative GAs with what they know about absolute GAs and control adjectives not presented as gradable. In addition, we offer a richer picture of the variable role of the context for the entire class of GAs.

Turning to definite descriptions, although knowledge of their presuppositions are clear for adults, the picture for children is murkier. Children's overreliance on the definite determiner in earlier production studies (cf. Karmiloff-Smith, 1979; Maratsos, 1976) has been interpreted as showing that they do not recognize the presuppositions of definiteness

associated with the lexical representation of the definite determiner (Wexler, 2003), while their performance in eyetracking experiments (cf. Trueswell, Sekerina, Hill, and Logrip, 1999) has suggested that they experience difficulty parsing definite DPs in real time when postnominal information must also be integrated. However, both of these findings actually leave unresolved the more basic question of whether children are *aware* of the presuppositions that a singular definite DP carries.

Showing that children do not produce the definite determiner when it is felicitous to do so is more revealing about how aspects of the context at hand motivate the utterances they *produce*, and not how their evaluation of utterances intended to capture aspects of the context reflects their semantic *competence*. Likewise, online parsing difficulties indicating that it may be challenging for children to incorporate other contextual information to restrict the reference of the noun (cf. Hurewitz *et al*. 2000; Meroni, 2006) may mask children's underlying knowledge of the presuppositions associated with definiteness. By using definite descriptions to draw out the distinctions between relative and absolute GAs, we not only demonstrate that children have different representations for these GAs (one being context-dependent and the other not), but that children appreciate that the presuppositions of this class of expressions.

2.4 Overview of the Experiments and Main Results

In the remainder of this paper, we present a set of experiments involving differentiation uses of definite descriptions with GAs. The results demonstrate that children as young age three and adults assign context-dependent interpretations only to relative GAs, and that they are sensitive to – and can correctly accommodate – the presuppositions of definite DPs. We then highlight an important difference between children and adults involving the maximum standard absolute GAs *full* and *straight.* Given a request with a definite DP of the form *the full/straight one* in a context in which neither of the two objects

exhibits the relevant property to a maximum degree, adults uniformly reject this request, illustrating their implicit acknowledgement of the failure of the existence presupposition, while children systematically accept such uses. Crucially, though, children are less likely to do this if they had already seen the maximal standard and take significantly longer to accept such requests than those with relative GAs or true predications of these GAs. We argue that this finding supports a distinction between the true semantic context dependence (vagueness) of relative GAs on the one hand, and contextually acceptable degrees of imprecision associated with maximal standards on the other.

Our study has three major results. First, it gives the earliest experimentally identified age at which children have achieved adult-like levels of competence in the fundamental aspects of the context/meaning relation we target (three years old). Second, it offers experimental support from both child and adult participants for a semantic distinction between relative and absolute gradable adjectives based on the (non-)context dependence of the standard of comparison. Third, it supports a distinction between the two types of contextual variability discussed at the end of section 1.3: one that is fundamentally *semantic* in nature (variability in the extension of relative GAs) and one that is fundamentally *pragmatic* (tolerance for imprecise uses of absolute GAs).

3. Scalar Judgment Task

In previous work (Syrett, Bradley, Kennedy, and Lidz, 2006) we examined the extent to which young children's and adults' judgments along a scale reflect distinctions among the three different subclasses of GAs: relative GAs, absolute maximum standard GAs, and minimum standard GAs.[8] By targeting representative exemplars from each class (*big*, *long*;

[8] We tested 36 three-, four-, and five-year-olds' (12 children in each age group) and 28 adult controls. Further details regarding age, materials, experimental methodology, and results are provided in Syrett, Bradley, Kennedy, and Lidz (2006).

full, spotted, respectively), we sought to provide experimental evidence for the division among these classes that would motivate the experiments outlined below. We presented each participant with a set of seven items for each target GA (blocks, rods, containers, discs) and one at a time, starting at the positive polar end of the scale[9], we simply asked of each item in the set, *Is this A?*, where A was the target GA corresponding to the target property.

What we found was a clear division among GAs. For both age groups, there was a decrease in the acceptance for both relative GAs (*big*, *long*) around the midpoint of the series. This is to be expected if the standard of comparison shifts given the context at hand. Judgments for the absolute GAs, however, did not show this pattern. Both children and adults accepted as *spotted* any disk with any number of spots on it, rejecting only the disk with no spots. Judgments for the maximum standard absolute GA *full* were noticeably different from the minimum standard absolute GA *spotted*: adults only accepted the first container as *full*, and children's acceptance of what counted as *full* began to drop sharply beginning with the second item in the series, differing significantly from both *spotted* and the two relative GAs *big* and *long*. Children were also significantly less likely to judge the second container to be full if they began with this set of objects, instead of with one of the two relative GA sets, an

[9] In Experiments 1 and 2, we refer to the item at this end as #1, and the item at the opposite end as #7. Throughout the experiments reported in this paper, we make use of the increasing/polar-positive member of an antonym pair (e.g., *big*, not *small*, *full*, not *empty*), since our focus was on context sensitivity and not on the asymmetry between poles. Certainly a sizable literature has been devoted to investigating this phenomenon in child language (cf. Barner and Snedeker, 2007, 2008; Bartlett, 1976; Brewer and Stone, 1975; Carey, 1978; Clark, 1972; Clark, 1973; Eilers, Kimbrough Oller, and Ellington, 1974; Keil and Carroll, 1980; Klatzky, Clark, and Macken, 1973; Marschark, 1977; Townsend, 1976).

effect that makes sense in the context of the ordering effect discussed in the results section of the following experiments.

The Scalar Judgment Task (SJT) thus reveals an important difference in the way the standard of comparison is set among these GAs. Relative GAs such as *big* and *long* have standards that evoke a partitioning of a finite set of objects ranked along a scalar continuum somewhere around the midpoint of the continuum. Absolute GAs such as *spotted* and *full* use standards that partition the sets at the ends of the continuum – *spotted* distinguishing the lowest-ranked item (the only object lacking the relevant property) from the rest of the series and *full* distinguishing the highest-ranked item (the object maximally possessing the relevant property) from the rest of the series. This set of findings sets the stage for the following experiments, where we further probe distinctions among these GAs related to the role of context and the way in which the standard of comparison is set.

4. Experiment 1

4.1 Experiment 1a

4.1.1 Introduction

In Experiments 1 and 2, we used definite descriptions to probe children's sensitivity to the kinds of context/meaning interactions discussed in the introduction. Specifically, we were interested in determining whether children would correctly shift the standard of comparison for relative GAs to accommodate the existence and uniqueness presuppositions of the definite description, and would correctly avoid doing so for absolute GAs.

4.1.2 Method

Participants

Thirty children representing three age groups participated in this task: 10 three-year-olds (5 boys 5 girls, range: 3;5 to 3;11, M: 3;8); 10 four-year-olds (4 boys 6 girls, range: 4;1 to 4;11, M: 4;5); and 10 five-year-olds (3 boys 7 girls, range: 5;1 to 5;8, M: 5;5). 24 adults

served as controls. All adults in our experiments were Northwestern undergraduates fulfilling an experimental requirement for a Linguistics course and were native speakers of English.

Materials

The materials consisted of a series of pairs of objects, each sharing a salient dimension (e.g., color, shape, length, etc.). The experiment was divided into a training session and test session. The complete set of materials is presented in the Procedure section.

Procedure

Participants were invited to play a game. Children were introduced to a puppet (played by a second experimenter) and were told that that the purpose of the game was to help the puppet learn how to ask for things. They were then told that they would be shown two objects at a time and that every time they saw two objects, the puppet would ask for something. Their job was to determine if they could give the puppet what he asked for based on his request, and if they couldn't, to tell him why not. Even the youngest participants followed these directions easily and were eager to participate. Adult participants interacted with an adult experimenter instead of the puppet.

There was an important twist in this task that distinguished it from previous forced-choice studies in which a child was asked to act on a request with an adjective such as *big* (e.g., Bartlett, 1976; Ebeling and Gelman, 1988; Eilers, Kimbrough Oller, and Ellington, 1974; Gelman and Ebeling, 1989; Gelman and Markman, 1985; Harris, Morris, and Terwogt 1986; Ravn and Gelman, 1984; Sena and Smith, 1990). In our task, the request was not always felicitous. This pragmatic manipulation was accomplished in the following way. Each request included a singular definite DP of the form *the A one*, where *A* was an adjective (e.g., *Please give me <u>the red one</u>*). As noted in section 1.2, this type of description presupposes both existence (e.g., there must *be* a red one) and uniqueness (e.g., there must be *only one* red one). We deliberately varied the felicity conditions of the request, presenting participants

with pairs of objects that they either satisfied or violated one or both of the presuppositions of the definite description. Therefore, for some pairs, the request (e.g., *Give me the red one*) was *felicitous*, because exactly one object fit the description (e.g., there was one red and one white object). While for other pairs, the request was *infelicitous*, either because *both* members of the pair fit the description (e.g., there were two red objects) or *neither* did (e.g., there was a yellow object and a blue object). In determining whether or not they could give the puppet what he asked for, children were in essence, assessing the context with respect to the presuppositions of the definite description, and in some cases accommodating them. For this reason, we refer to this task as the *Presupposition Assessment Task (PAT)*.

To ensure that participants understood the rules of the game and to help them feel comfortable rejecting the request, we had them participate in a brief training session before the test session began. This training session was composed of two felicitous and two infelicitous requests with the adjectives *happy*, *round*, *red*, and *blue*, similar in design to the control items in the test session. If children were still hesitant to correct the puppet after the four training items, we introduced a fifth impromptu pair accompanied by an infelicitous request. Once it was evident that participants felt comfortable with the task, we proceeded with the test session.

Items in the test session included target items whose salient property corresponded to one of the target GAs, as well as control items. Descriptions of the stimuli are presented in the following two tables. Table 1 presents the control stimuli used in the test session, while Table 2 presents the target stimuli used in the test session. The target GA items were taken from the four sets in the SJT. This allowed us to compare baseline judgments of the stimuli to how participants judged them in the PAT. One pair in each of the two relative GA sets consisted of objects that had been judged to possess the relevant property, while the other pair consisted of objects that were judged not to have that property.

The presentation of the pair members was counterbalanced so that the object fitting the description appeared on different sides of the pairs throughout the test session. The order of the pairs was also pseudorandomized with respect to three factors: the felicity of the request, the nature of the presupposition violation, and the adjective. Participants were randomly assigned to one of two orders of presentation, a point that becomes important in the interpretation of the results.

Table 1: Control stimuli for Experiment 1

adjective	stimuli	pragmatic status of request
color		
yellow	pictures of a yellow bird and a black bird	felicitous
red	red poker chip and a white poker chip	felicitous
green	purple yo-yo and yellow yo-yo	infelicitous
red	pictures of a red square and a red circle	infelicitous
shape		
square	pictures of a blue square and a yellow circle	felicitous
round	pictures of a red triangle and a red square	infelicitous
mood		
sad	pictures of a sad face and a happy face	felicitous
happy	pictures of a sad face and an angry face	infelicitous
happy	pictures of a sad face and an angry face	infelicitous

Table 2: Target stimuli for Experiment 1

adjective	stimuli	pragmatic status of request
relative		
big	two blocks from "upper" end of SJT sequence (the ones judged big), one bigger than the other	felicitous
big	two blocks from "lower" end of SJT sequence (the ones judged not big), one bigger than the other	felicitous
long	two rods from upper end of SJT, one longer than the other	felicitous
long	two rods from lower end of SJT, one longer than the other	felicitous
absolute		
spotted	two disks, one with some spots, one with none	felicitous
spotted	two disks, one with some spots, one with more	infelicitous
full	two jars, one full, one about 2/3 full	felicitous
full	two jars, neither full but one fuller than the other	infelicitous

Predictions

We based our predictions for each of the four target GAs on the semantic GA classification they represent. If indeed only relative GAs depend on the context to set the standard of comparison and therefore allow the standard to shift from context to context, then we should only see accommodation of the presuppositions of the definite description with the relative GAs *big* and *long*. Thus if a new standard of comparison is posited each time a new pair is introduced in order to ensure that the adjective is be true of just one object (i.e., the bigger or longer one), the request should always be felicitous and participants should always

accept the request. (See the discussion of this point in section 1.2.) This pattern of responses would constitute evidence that children know that such GAs are context-sensitive in the way outlined in section 1.2, and that they are sensitive to, and willing to accommodate, the presuppositions of a definite DP. If, however, judgments about the size or length of the items remains constant between the SJT and the PAT, without the standard shifting, then either an existence or uniqueness violation should be incurred, depending on whether the two objects are from the negative or positive polar end of the continuum, respectively. For example, two blocks otherwise judged to be *big* should incur a uniqueness violation, while two blocks that were judged to be *not big* should incur an existence violation, unless the standard shifts.

Pairs involving absolute GAs provide important controls and substantiation of this conclusion, as outlined in section 2.2. If absolute GAs in the positive form have fixed (maximum or minimum) standards of comparison and are not context-dependent, they should not allow the same flexibility of use as relative GAs. For example, if *spotted* simply means 'has some number of spots', participants should reject requests for *the spotted one* when confronted with two spotted objects, even if one is clearly more spotted than the other, because this would incur a uniqueness violation, as both objects are spotted. Likewise, if *full* means 'is maximally full', participants should reject requests for *the full one* when confronted with two partially filled containers, even when one is closer to full than the other, because the request incurs an existence violation, as there is no full container.

If, on the other hand, *spotted* and *full* are treated as context-dependent, participants ought to be able to shift their meanings in a way that accommodates the presuppositions of the definite description in these examples. Thus, rejection of presupposition-violating requests with the absolute GA pairs, coupled with acceptance of requests with the relative GA pairs, would constitute evidence for a distinction among these GAs that would appear to be based on the relative/absolute GA distinction laid out earlier and would show that

participants are, in fact, modulating the interpretation of these relative GAs in accord with the presuppositions of the definite DP.

At the same time, this pattern of results would also provide important evidence that participants are not just treating the positive form of the adjective in the definite description as semantically equivalent to the comparative form, treating *Please give me the A one* as *Please give me the more A one*. If participants were reinterpreting the request in this way, they should *always* accept it, regardless of adjective type, since the comparative form of any GA can be used to uniquely pick out that member of a pair that has the greater degree of the relevant property: *the more spotted one* can be felicitously used to pick out the more spotted of two spotted disks, even though *the spotted one* cannot. Although it is unlikely that adults would reinterpret *the A one* in this way, it is a possibility that must be seriously considered for children, as their interpretation and use of comparative morphology at this young age is not fully adult-like (cf. Donaldson and Wales, 1970; Ehri, 1976; Finch-Williams, 1981; Gathercole, 1979; Gitterman and Johnston, 1983; Graziano-King, 1999; Graziano-King and Cairns, 2005; Layton and Stick, 1978; Moore, 1999). If participants reject presupposition-violating requests involving absolute GAs, however, we can be confident that they are not simply reanalyzing the adjectives as semantically equivalent to their comparative counterparts.

4.1.3 Results

Recall that for each pair, we predicted one of three possible responses to the accompanying request, based on the above felicity patterns: participants would accept the request and give one object, give neither, or (perhaps rather reluctantly) give both objects. Both the second and third responses count as rejecting the request, since both highlight a presupposition failure of the singular definite description. Note that whenever one object is given, such a response can only be considered appropriate when the member of the pair with

the greater degree of the relevant property is targeted, since no matter how the request is interpreted, the adjective can never be interpreted as highlighting a degree below the cutoff. Thus, we recorded which of the two objects the participant gave in response to the request – the member with the greater or lesser degree. The distribution of responses is presented in Table 3.

Table 3: Distribution of responses in Experiment 1a

		age group											
		3 yrs			4 yrs			5 yrs			adults		
		1+,	1-,	0/2	1+,	1-,	0/2	1+,	1-,	0/2	1+,	1-,	0/2
1	control-felicitous	98,	0,	2	100,	0,	0	100,	0,	0	100,	0,	0
2	control-infelicitous	4,	6,	90	0,	2,	98	0,	0,	100	0,	0,	100
3	*big* (upper SJT)	100,	0,	0	100,	0,	0	100,	0,	0	100,	0,	0
4	*big* (lower SJT)	90,	0,	10	90,	0,	10	100,	0,	0	96,	0,	4
5	*long* (upper SJT)	80,	0,	20	90,	0,	10	100,	0,	0	100,	0,	0
6	*long* (lower SJT)	100,	0,	0	100,	0,	0	100,	0,	0	96,	0,	4
7	*spotted* (some spots/more spots)	20,	0,	80	30,	0,	70	10,	0,	90	4,	0,	96
8	*spotted* (some spots/no spots)	100,	0,	0	100,	0,	0	100,	0,	0	100,	0,	0
9	*full* (full/less full)	100,	0,	0	100,	0,	0	100,	0,	0	96,	0,	4
10	*full* (not full/less full)	60,	0,	40	70,	0,	30	70,	0,	30	12,	0,	88

Three clear trends stand out in these data. First, all participants regardless of age group gave the ‘greater degree’ member of the pair in response to felicitous requests in which

only one of the two objects satisfied the request (the control items in row 1 and the *full* and *spotted* test items in rows 8 and 9). Indeed, in all but one trial, across all ages, participants consistently gave the one item with the relevant property. Second, participants consistently rejected infelicitous requests for the control items (row 2) and the 'some spots/more spots' test items (row 7). For example, there is no difference between the target *spotted* pair and the control *shape* pair, both of which participants saw once (Fisher's exact test, children: $p > 0{,}22$, adults: $p = 1.00$). These results provide us with a baseline against which we can evaluate participants' responses to requests using relative GAs. Here, a third trend emerges: despite judgments of individual items in the SJT, participants across age groups also almost always gave the object that had the greater degree of the relevant property when asked for the *big* or *long* one (cf. rows 3-6). We found no significant difference in responses to the trials containing relative adjectives and the control trials (Pearson's $\chi^2 = 0.036$, $df = 2$, $p > 0.98$). We take this as an indication that participants were willing and able to shift the standard of comparison for the two relative GAs and therefore that they have acquired adult-like context-dependent denotations for these adjectives.

The one instance in which children's responses diverge from adults' is with the *full* pair in which neither container is full (row 10) (Pearson's $\chi^2 = 13.14$ (with Yates' continuity correction), $df = 1$, $p < 0.001$). Instead of rejecting the puppet's request when shown the 'not full/less full' pair (as adults categorically did), children gave the puppet the fuller of the two containers. However, this cannot be because the children actually treated all of the adjectives as semantically comparative (understanding *the full one* as *the fuller one*): if this were an option, they should have applied the same strategy to *spotted*, returning the more spotted of the two disks when shown the 'some spots/more spots' pair, in which case rows 7 and 10 of Table 3 would look the same, contrary to our observations (Pearson's $\chi^2 = 11.47$, $df = 1$, $p < 0.001$).

We also have reason to believe that many of these children knew that the fuller member of the pair was, in fact, not full. 18 children who participated in this experiment also participated in the SJT within approximately three weeks time. 12 (or 67%) of these children judged only the first container in the series to be full; four (22%) judged the first and second containers to be full; 2 (11%) judged the first three containers to be full; but no one judged containers beyond the third to be full. The fuller member of the pair in the 'not full/less full' condition was the fourth item in the series, so it was among the set of containers that were uniformly judged to be not full in the SJT. But in the PAT, 11 of the 18 children who had made this judgment (61%), across all age groups, returned this container in response to a request from the puppet for *the full one*.

So the question is whether children differ from adults in treating *full* as relative, or whether there is another explanation for their behavior. To answer this question, we decided to probe children's responses to this *full* pair further by examining whether the order to which participants had been assigned had had an effect. Indeed, upon further examination, we found that every single one of the 15 children assigned to the order in which this pair appeared earlier in the sequence than the pair that had the full container gave the puppet the fuller of the two containers in response to his request for *the full one*. Only five of the 15 children in the condition in which this pair appeared later in the sequence after the 'full/less full' pair responded to the puppet's request in this way (a difference of 100% and 33.3% between the two conditions). We address the implications of this findings in the following discussion section.

4.1.4 Discussion

Experiment 1a demonstrates that both adults and children distinguish between the target GAs in the way that the standard of comparison is assigned. Because relative GAs such as *big* and *long* have denotations that invoke a contextually-determined standard of

comparison, their extensions may shift from context to context. Accordingly, when presented with requests for *the big/long one*, participants readily gave the bigger or longer member of the pair. In contrast, when presented with two spotted objects, participants did not give the more spotted object; this is to be expected, since *spotted* is a minimum standard absolute GA with a meaning along the lines of 'has spots'. Adult participants similarly rejected requests for *the full one* when given two non-full containers, an expected response if *full* means 'maximally full'. However, many children gave the puppet the fuller container in this situation. This pattern contrasted with our findings from the SJT, in which we found that many of these same children did not judge this container to be full.

Post-hoc analysis of the data indicated that a relative pair appearing earlier in the sequence affected children's responses to the 'not full/less full' pair, increasing the probability of a 'relative-like' reaction. One potential explanation for this behavior is that, despite our findings in the SJT, children are uncertain as to whether *full* is absolute, but settle on such an interpretation (in the context of the experimental task) when they receive early exposure to an object that exemplifies the maximal standard (the full container). A second possibility is that children were influenced by their judgments about the examples involving the relative adjectives. In this experiment, whenever the 'not full/less full' pair appeared first, it was also immediately preceded by one of the pairs involving the relative adjective *long*. It is therefore possible that the relative adjective induced a kind of priming effect, causing the children to treat *full* as relative on analogy to their previous decisions about *long*. A second experiment was run to decide between these two possibilities.

4.2 Experiment 1b

4.2.1 Introduction

The goal of Experiment 1b was to determine the source of the order effect observed in Experiment 1a. Specifically, we sought to determine whether the prior presentation of a

relative GA pair in the sequence influenced children to treat *full* as context-dependent, causing them to give the puppet the fuller of the two 'not full/less full' containers in response to his request for *the full one*.

4.2.2 Method

Participants

Seventeen children representing three age groups also participated in this task: six three-year-olds (3 boys 3 girls, range: 3;1 to 3;11, M: 3;5); six four-year-olds (2 boys 4 girls, range: 4;2 to 4;11, M: 4;6); and five five-year-olds (1 boy 4 girls, range: 5;2 to 5;10, M: 5;4). 10 adults served as controls.

Materials

The same objects from Experiment 1a were used. The only difference was in the sequence of items. In Experiment 1a, the 'not full/less full' pair was almost immediately preceded by a *long* pair, with only one control pair intervening. To evaluate the influence of the relative GA pair, we simply switched the order of the *long* and 'not full/less full' pair so that the latter preceded the former.

Procedure

The procedure was the same as in Experiment 1a.

4.2.3 Results

Because our purpose in conducting Experiment 1b was solely to investigate the order effect described at the end of Experiment 1a, we only present responses to the two pairs whose position in the sequence was swapped (the *long* and 'not full/ less full' pairs) and the complementary 'full/less full' pair for comparison. (See Table 4.) All other responses were a replication of Experiment 1a.

Table 4: Distribution of key responses in Experiment 1b

	3 yrs	4 yrs	5 yrs	adults
	1+, 1-, 0/2	1+, 1-, 0/2	1+, 1-, 0/2	1+, 1-, 0/2
long (upper SJT)	100, 0, 0	100, 0, 0	100, 0, 0	100, 0, 0
full (full/less full)	100, 0, 0	100, 0, 0	100, 0, 0	100, 0, 0
full (not full/less full)	100, 0, 0	83, 0, 17	60, 0, 40	30, 0, 70

The pattern of responses is highly similar to the one observed in Experiment 1a, with children across age groups – but not adults – targeting the fuller 'not full/less full' container as *the full one* in response to the speaker's request (Pearson's $\chi^2 = 5.33$ (with Yates' continuity correction), $df = 2$, $p < 0.02$).[10] Thus, the order of the relative and target *full* pair was not responsible for the original pattern of responses.

A comparison of children's responses to the 'not full/less full' pair in Experiments 1a and 1b supports this conclusion. (See Table 5.)

[10] Adults who gave the fuller of these two containers noted at the end of the experimental session without any prompting that they realized their mistake later in the experiment and wished to make clear to the experimenter that they knew what *full* means.

Table 5: Percent of time participants rejected the request when shown the 'not full/less full' pair

	1	2	3
age	Experiment 1a: long < not full/less full < full/less full	Experiment 1b: not full/less full < long < full/less full	Experiment 1a: full/less full < not full/less full
children	0	18	67
adults	75	70	100

The difference in judgments between the two sequences that varied the order of the relative pair and the 'not full/less full' pair (a comparison between columns 1 and 2) is not significant (Fisher's Exact Test, two-tailed probabilities: 'column 1' v. 'column 2': $p = 0.24$); but the difference in judgments between the orderings in which the 'not full/less full' pair preceded the 'full/less full' pair and the order in which the 'full/less full' pair came first (columns 1, 2 v. column 3) is significant. ('column 1' v. 'column 3': $p < 0.001$; 'column 2' v. 'column 3': $p = 0.01$). The conclusion to be made here is that children were not misled in their interpretation of *full* by the prior presentation of a relative GA. Instead, we may conclude that their (un)willingness to treat the fuller of two the containers that were both not full as full is dependent on whether or not they have already seen an instance of maximal fullness.

That said, it remains the case that even when children saw the 'full/less full' pair first (column 3 in Table 5), their responses to the 'not full/less full' pair deviated significantly from those of adults. This is surprising considering the similarity in the two age groups' performance with all of the other control and target pairs, and also considering their earlier responses in the SJT, which suggested that they had acquired the absolute denotation for *full*. Taking these observations as a starting point, we hypothesized that their responses to the 'not full/less full' pair might indicate not a willingness to treat *full* as relative, but rather a

willingness to tolerate a certain amount of imprecision on the part of the puppet: a willingness to respond to a request for *the full one* in a context in which neither of the two objects uniquely satisfied the description by handing over the object that came closest to doing so. We further reasoned that if this were correct, it would imply some additional reasoning about the utterance beyond the computation of its semantic content: minimally, a recognition that the description used was false of the two objects, a determination of which object came closest to making it true, and a calculation of whether to behave as though it were true.[11]

To explore this possibility, we examined children's reaction times (RTs) during the experimental session. Experimental sessions with child participants were videotaped using a Sony Digital8 Handycam. Videotapes were imported from the camera onto a Macintosh computer as .mov files, which were then coded offline by research assistants in our laboratory using SuperCoder software (Hollich, 2003).[12] For each item in which the child accepted the puppet's request by giving the puppet one of the two objects, we coded three measurements: the child's *look* to this object, his/her subsequent *reach* toward this object, and his/her ultimate *touch* of this object. The videos were coded frame by frame, where one frame is equal to 1/30 of a second.

[11] We know from the ordering effects in Table 5, and from the systematic rejection of the 'some spots/more spots' pair, that children do not indiscriminately behave as though the description is uniquely true of one of the members of the pair. This implies that when they do behave in such a way, it involves some sort of reasoning on their part.

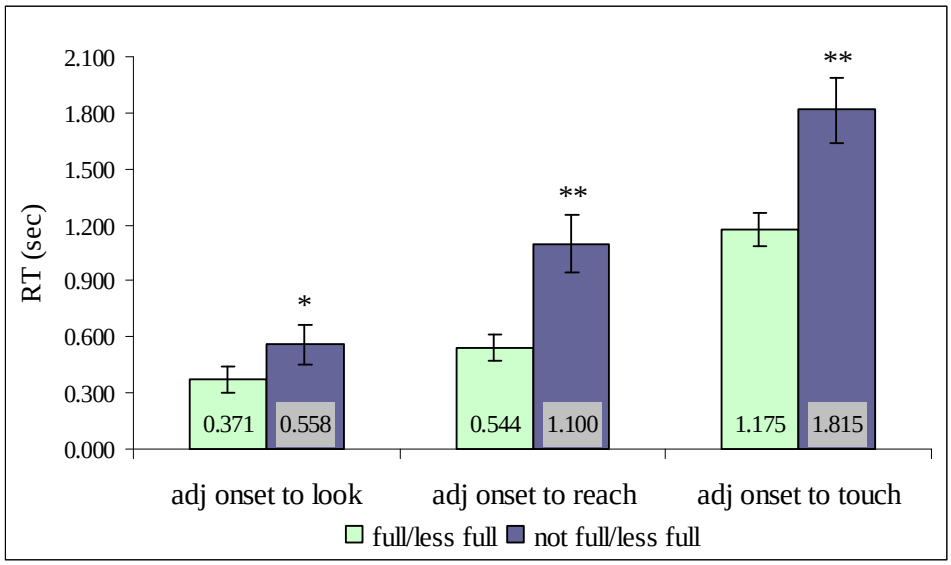
[12] At least two coders were assigned to each experimental session, with one coder arbitrarily chosen as the default. The inter-coder rate of agreement across all trials averaged above 95%. In case of disagreements of more than 5 frames for any of the three measurements, a third coder was brought in as a tiebreaker for that item.

It is necessary to provide two additional details about the coding session. First, we excluded from analysis any items in which the children's eye movements could not be coded (e.g., if the eyes were occluded), items in which the child was already looking at or touching the stimuli before the request was uttered, and items in which any other experimental artifact prevented the coders from obtaining measurements (e.g., there was a distraction in the background). For this reason, the total number of children whose RTs were analyzed varies from analysis to analysis. This number is always provided in a footnote. Second, rather than coding the initial look to the object, since the child could have decided to inspect the second object before deciding to give the puppet the first object, we coded the look that immediately preceded the reach to the object. A reach was a movement that ultimately resulted in touching the object. We chose to target these measurements instead of proportion of looking time, since we wished to measure latency of response. We targeted two sets of RT measurements for analysis, looking across the two experiments that made up Experiment 1.

We looked first at just the two *full* pairs, asking whether children took longer to respond to the puppet's request when it accompanied the 'not full/less full' pair than when it accompanied the 'full/less full' pair. These results are presented in Figure 1.[13] Indeed, differences between the look, reach, and touch are significant for these two pairs (one-tailed t tests: look $t(15) = 1.71$, $p = 0.05$; reach $t(15) = 3.03$, $p = 0.004$; touch $t(15) = 3.47$, $p < 0.002$).

[13] We analyzed the RTs for 16 children who targeted the fuller container for each pair.

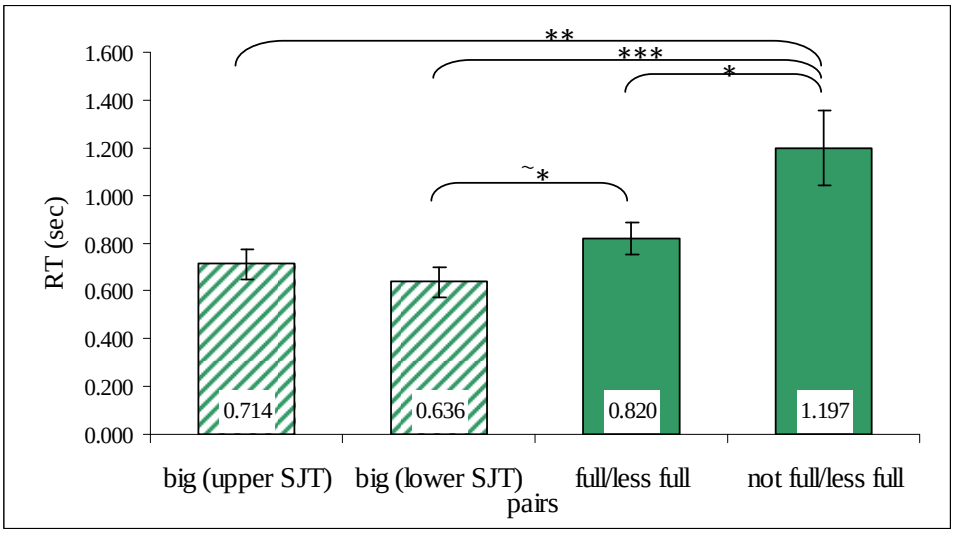
Figure 1: Reaction times for two *full* pairs in Experiment 1a



Second, we looked across GA subclasses and compared the difference between the look and the touch of the object for the 'not full/less full' pair to those for three other key pairs – the two *big* pairs and the 'full/less full' pair – in order to determine if the increase in RTs was unique to the 'not full/less full' pair. These results are presented in Figure 2.[14]

[14] We analyzed responses to these pairs across all children. The number of children varied for each pair: 26 for the blocks from the positive pole, 32 for the blocks from the negative pole, 29 for the 'full/less full' pair, and 21 for the 'not full/less full' pair. We could not compare performance with the *spotted* pairs, given the small percentage of cases in which children actually gave the puppet the more spotted object.

Figure 2: Reaction times for four key pairs in Experiment 1a



The RTs for the 'not full/less full' pair are significantly longer than every other pair (two-tailed t-tests: 'not full/less full' v. 'full/less full': $t(48) = 2.42$, $p = 0.02$; 'not full/less full' v. 'big (upper SJT): $t(45) = 3.07$, $p < 0.01$; 'not full/less full' v. 'big (lower SJT)': $t(51) = 3.79$, $p < 0.001$), while the RTs for the other pairs do not differ significantly from each other (two-tailed t-tests: 'full/less full' v. big-positive: $t(53) = 1.11$, $p = 0.27$; 'big (upper SJT)' v. 'big (lower SJT)': $t(56) = 0.87$, $p = 0.39$; 'full/less full' v. 'big (lower SJT)': $t(59) = 1.96$, $p = 0.05$, marginally significant).

What these RTs show is that in the case of the relative GA *big*, a shift in standard of comparison is automatic, at least in the context of accommodating the presupposition of a definite description. Furthermore, this holds in both directions: a shift *up* to distinguish between two objects from the upper end of the SJT sequence (rendering *big* false of one of two blocks previously judged to be 'big'), and a shift *down* to distinguish between two objects from the lower end of the SJT sequence (rendering *big* true of one of two blocks previously judged to be 'not big'). By contrast, RTs increased when children gave the puppet a container that was not full in response to his request for *the full one*, as we hypothesized. We assess the theoretical significance of this pattern in Section 6.

4.2.4 Discussion

The results of Experiment 1b elaborate upon those of Experiment 1a by demonstrating that children's decision to give the puppet the fuller of the 'not full/less full' containers in response to his request for *the full one* is not driven by the preceding *long* relative GA pair. Regardless of the order of these two target pairs, children are likely to accept the puppet's request and give him the fuller container if they have not already seen the 'full/less full' pair, which exemplifies the maximum standard. However, even with this assistance, some children are still inclined to return the fuller of the 'not full/less full' pair when asked for *the full one*.

Our RT analysis shed light on this pattern by demonstrating that children take significantly longer to give the puppet the fuller of the two containers that are not full than they take to return a container that is actually full or the larger member of the pairs in the two *big* conditions, regardless of the baseline judgments for these objects. This difference in RTs suggests that there is another layer of processing associated with their behavior on the crucial items. We provide a more detailed analysis of this phenomenon in the Section 6. Before turning to this, we describe a second set of experiments involving different absolute adjectives that were designed to ensure that our results in Experiments 1a-b generalize beyond *full* and *spotted*.

5. Experiment 2

5.1 Introduction

Given the nature of children's responses to the *full* pairs in Experiment 1, one question immediately surfaces: are these responses unique to *full*, or can they also be observed with other maximum standard absolute GAs? The goal of Experiment 2 was to address this question. We also introduced another absolute minimum standard GA, *bumpy*, so that we could generalize our results to two instances in each of the three GA subclasses (relative, absolute maximum standard, and absolute minimum standard).

5.1.2 Method

Participants

Thirty children representing three age groups participated in this task: 10 three-year-olds (3 boys 7 girls, range: 3;2 to 3;10, M: 3;6); 10 four-year-olds (4 boys 6 girls, range: 4;1 to 4;9, M: 4;7); and 10 five-year-olds (5 boys 5 girls, range: 5;1 to 5;11, M: 5;6). 24 adults served as controls.

Materials

The materials were the same as in Experiment 1, with the exception of the absolute GA pairs. In place of the two *full* pairs, there were two pairs corresponding to the maximum standard absolute GA *straight*.[15] For example, a straight wire was paired with one that was straight for most of its length but which had a curl at the top, analogous to a container which is filled most, but not all, of the way.[16]

In place of the two *spotted* pairs, there were two pairs corresponding to the minimum standard absolute GA, *bumpy*. These pairs were designed similarly to those in Experiment 1, so that only one of the two pairs for each adjective would satisfy the presuppositions of the definite description. (See Table 6.)

Table 6: Absolute GA stimuli used in Experiment 2

---

[15] The fact that *full* and *straight* are both absolute maximum standard absolute GAs can be highlighted by adverbial modification. While both of these adjectives may be modified by either *completely* or *perfectly*, the relative GAs *big* and *long* cannot. Likewise, the fact that *spotted* and *bumpy* are both minimum standard absolute GAs can be highlighted by modifying them by *slightly,* an adverb that cannot appear with any of the other target GAs.

[16] See Syrett, Lidz, and Kennedy (2006a, b) for further discussion of the *straight* stimuli selected for this task.

adjective	status of request	stimuli
straight	felicitous	a straight wire and a straight wire with a curly section at the top 
	infelicitous	a straight wire with a curly section at the top and a totally curly wire 
bumpy	felicitous	two orange wooden 2″ x 5″ boards, one with some bumps, one with none
	infelicitous	two orange wooden 2″ x 5″ boards, one some bumps, one with more

Procedure

The procedure was the same as in Experiment 1.

5.1.3 Results

The results for Experiment 2 are presented in Table 7. These results can be compared with those from Experiments 1a and 1b, presented in Table 3 and Table 4, respectively. As the key comparison between experiments is between the two sets of absolute GA pairs, we only present participants' responses to these pairs in Table 7. (All other responses were replications of Experiment 1.)

Table 7: Distribution of responses in Experiment 2

	3 yrs	4 yrs	5 yrs	adults
	1+, 1-, 0/2	1+, 1-, 0/2	1+, 1-, 0/2	1+, 1-, 0/2
bumpy (some bumps, more bumps)	20, 10, 70	20, 10, 70	10, 10, 80	0, 0, 100
bumpy (some bumps, no bumps)	70, 30, 0	90, 10, 0	100, 0, 0	96, 0, 4
straight (straight, bent)	90, 0, 10	90, 10, 0	90, 0, 10	100, 0, 0
straight (bent, more bent)	70, 0, 30	40, 0, 60	20, 10, 70	12, 0, 88

Children's responses are similar to those seen in Experiment 1 with the exception of behavior of the 4- and 5-year-olds in response to the 'bent/more bent' pair; these children were more likely to accept such requests with the 'not full/less full' pair (Pearson's $\chi^2 = 4.9$ (with Yates' continuity correction), $df = 1$, $p < 0.03$). Once again, however, we observed an ordering effect whereby children were more likely to accept the puppet's request for *the straight one* if they had not yet seen the maximal standard: nine of the 13 children (69%) who gave the puppet the straighter of the two bent wires saw this pair appeared early in the sequence of items. A comparison of the two orders of presentation highlights the consistency of this effect across the three experiments. (See Table 8.)

Table 8: Percent of time children rejected the puppet's request for *the full/straight one* when shown a pair without a maximal standard

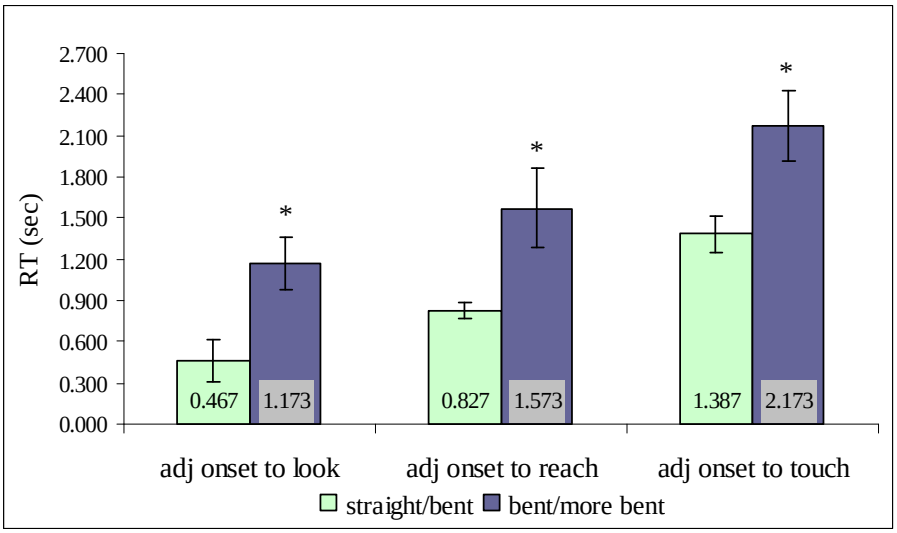
	Experiment 1a	Experiment 1b	Experiment 2
before maximal pair	0	18	38
after maximal pair	67	n/a	71

This across-the-board pattern allows us to make the following generalization: before encountering a true maximal standard, children are more likely to accept a request for an object exemplifying such a standard when they are shown a pair in which one object is close enough to this standard to count as such – even when they know it to be the case that this object does not actually exemplify the standard (as in the *full* examples). This pattern appears to reflect an allowance of a degree of imprecision in the extent to which the maximal standard holds.

In Experiment 1, we conducted an RT analysis, which revealed that children took longer to accept the puppet's request with the non-maximal *full* pair than when shown either the 'full/less full' pair or both *big* pairs. We now ask if the same trend holds for the *straight* pairs in Experiment 2. Although there were only five children for which the within-experiment comparison of the 'straight/bent' and 'bent/more bent' pairs could be made,[17] the same pattern emerges. (See Figure 3.) As with the *full* data, the differences between the adjective onset to the look, reach, and touch between the two pairs are all significant (one-tailed t tests: look $t(4) = -3.23$, $p = 0.016$; reach $t(4) = -3.15$, $p = 0.017$; touch $t(4) = -3.32$, $p = 0.015$).

[17] Recall that we were constrained with respect to which videos we could code, for reasons outlined earlier.

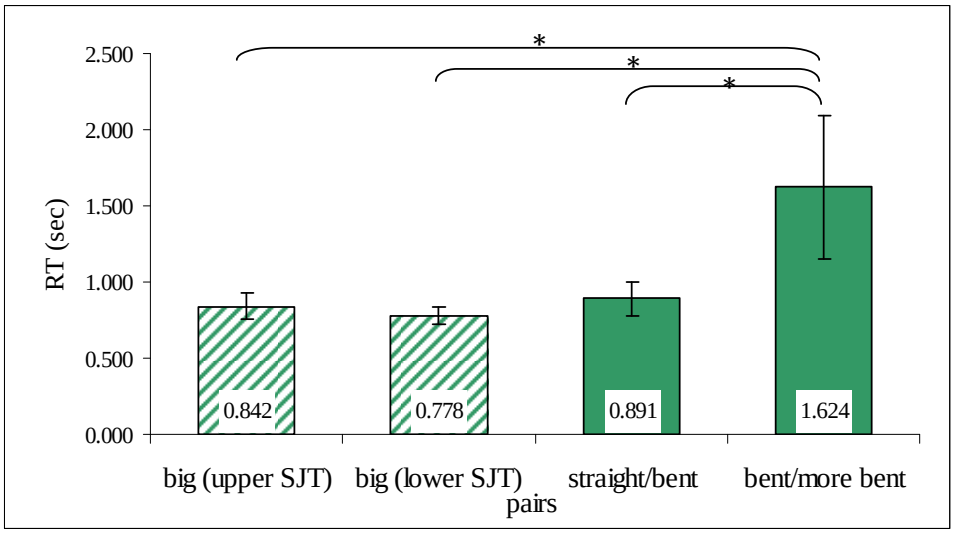
Figure 3: Reaction times for two *straight* pairs in Experiment 2b



As in Experiment 1, we also compared RTs across subjects, examining the difference between the look and the touch for the 'straight/bent' pair, the 'bent/more bent' pair, and the two *big* pairs. Again, a pattern similar to that seen in Experiment 1 emerges. (See Figure 4.) Children took significantly longer to select the straighter of the two bent pairs to satisfy the puppet's request than they did for the other three pairs (two-tailed t-tests: 'bent/more bent' v. 'straight/bent: $t(23) = 2.15$, $p = 0.04$; 'bent/more bent' v. 'big': $t(20) = 2.31$, $p = 0.03$; 'bent/more bent' v. 'big (lower SJT): $t(22) = 2.75$, $p = 0.01$). No other significant differences were found (two-tailed t-tests: 'straight/bent' v. 'big (upper SJT)': $t(31) = -0.33$, $p = 0.74$; 'big (upper SJT)' v. 'big (lower SJT)': $t(30) = 0.64$, $p = 0.53$; 'straight/bent' v. 'big (lower SJT)': $t(33) = 0.86$, $p = 0.40$).[18]

[18] The data from the following number of children were analyzed for each pair: 15 for the blocks from the positive pole, 17 for the blocks from the negative pole, 18 for the 'straight/bent' pair, and 7 for the 'bent/more bent' pair.

Figure 4: Reaction times for four key pairs in Experiment 2



5.2.4 Discussion

The results of Experiment 2, which replicate those of Experiment 1, support a generalization to two additional absolute GAs, *straight* and *bumpy*. Specifically, the effect of the order of presentation and the longer RTs for the non-maximal pair are not unique to the stimuli from in Experiment 1, and therefore not unique to the lexical item *full*. Rather, the fact that we see the same pattern with the absolute maximum standard GA *straight* indicates that the interpretive processes involved in choosing an object that comes closest to satisfying a description based on a maximum standard GA are different from those involved in choosing an object that satisfies a context-dependent description based on a relative GA. In short, understanding *the big/long one* as a description of the bigger/longer of two objects of unequal size is automatic, while tolerating use of *the full/straight one* as a description of the fuller/straighter of two objects when neither of them is actually *full* or *straight* requires some additional work.

6. General Discussion

The experiments presented here provide new insights into children's competence in three different aspects of the context/meaning relation. First, children's responses in the

Presupposition Assessment Task shows that by three years of age, they are not only aware of the existence and uniqueness presuppositions of a singular definite description, but that they are also willing to reject as infelicitous utterances that violate them. For example, in rejecting the puppet's request, children often offered statements such as the following: *Oh, but I have TWO red ones!*; *What red one? He should say what shape!*; *He thinks there must be two different colors!*; *They're ALL spotted!*; *The VERY bumpy one?*; *What bumpy one?* At the same time, their responses to the relative GA pairs show that when the semantics of the adjective and the context of utterance allow for the possibility of presupposition accommodation – in these cases, by shifting the standard of comparison so that the adjective (and, consequently, the description) is true of just one member of a pair of objects – children, like adults, will make the appropriate adjustment. Together, these results indicate that (at least in this domain) children are constructing the type of complex discourse models that linguistic expressions with presuppositions must be integrated into, and are willing and able to change those models to allow integration when such a move is licensed.

Second, we show that by three years of age, children correctly assign context-sensitive denotations to the relative GAs *big* and *long* when they are in the positive form: they are able to shift the standard of comparison for these adjectives in a way that is appropriate for the context of utterance. In the PAT, this shift was initiated in response to the pragmatic demands of the definite description in which the adjective appeared, and resulted in a change in the extension of the predicate. That children did not assign the same kinds of context-sensitive meanings to absolute GAs in the same contexts shows that they have already made subtle distinctions between predicates which are otherwise semantically quite similar: both relative and absolute GAs share the fundamental feature of expressing a relation

to a scalar concept, but they differ in whether the positive form denotes a relation to a context sensitive standard of comparison or to a fixed one.[19]

Finally, our results provide important new data on questions about the relation between context and (different kinds of) meaning. The results from the PAT show that children treated the relative GAs *big* and *long* and the maximum standard absolute GAs *full* and *straight* as 'variable'. For example, *the big one* was consistently accepted as a description of the bigger of two blocks, even when the blocks were otherwise judged not to be big, while *the full one* was often accepted as a description of the fuller of two containers, even when the container was not full. The latter result occurred less often and was influenced by the order of presentation; nevertheless it occurred enough to support the conclusion that some sort of interpretive variability is at work. The question is whether this variability is due to the same factors that are involved in interpretations of relative GAs – semantic context dependence – or whether it stems from a different source. It is important to emphasize that this is a much more general question, which bears on both child and adult language. While we did not observe the same variability in the maximum standard absolute GAs *full* and *straight* with adults, we suspect that we would have if the materials had been slightly different. In particular, we believe that if the fuller of the 'not full/less full' containers had been closer to

[19] According to Kennedy and McNally (2005) and Kennedy (2007), this difference actually stems from a more basic difference between the two classes of GAs, namely the structures of the scales that represent the gradable concepts they encode: relative GAs use open scales, while absolute GAs use closed scales. (See also Rotstein and Winter (2004).) If this is correct, then a more accurate way to state the generalization is that by three years of age, children already know the mapping between scale structure and the denotation of the positive form. See Syrett (2007) for extensive discussion of this issue.

full than it actually was (though still noticeably not full) adults would also have been inclined to provide it in response to the experimenter's request for *the full one*.

Our hypothesis is based on simple observation of everyday linguistic behavior: we regularly speak imprecisely. In addition to using absolute GAs like *full* and *straight* to describe objects that are not (strictly speaking) maximally full or straight, we might also describe our arrival somewhere at 3:03 p.m. by saying *We arrived at 3 p.m.*, or a child's consumption of all but a few bites of the food on her plate by saying *The child ate all her dinner*. (See Lasersohn (1999) and Sperber and Wilson (1986) for extensive discussion of such cases.) Given that our willingness to tolerate such loose talk is itself a matter of context (i.e., if the precise time of arrival is important, then the use of *arrived at 3 p.m.* to describe an arrival at 3:03 p.m. is inappropriate), it could be the case that like relative GAs, these expressions have context-dependent denotations. That is, it could be that interpretive variability is always fundamentally semantic, and that expressions like *full* and *straight* (as well as expressions such as *arrived at 3 p.m.*, *ate her dinner*, etc.) have meanings that, like *big* and *long*, require fixing the value of some contextual parameter before they can be assigned actual extensions. The denotation of an absolute GA in the positive form might be slightly different from a relative one – *full*, for example, might have a meaning along the lines of 'have a degree of fullness that is as close to maximally full as required by the contextual standard.' Importantly, though, on this view, any observed context dependence in relative and absolute GAs would be of fundamentally the same sort: fixing a parameter of evaluation in accord with the principles governing semantic composition.

This conclusion would be consistent with children's responses in the PAT (whether they gave the puppet an object to satisfy his request or not); to our minds, however, it is not consistent with the reaction time data that we collected. First consider the fact that the RTs for definite descriptions with the relative GA *big* were the same regardless of whether they

were presented with two objects from the positive pole (both objects were judged to be *big* in the Scalar Judgment Task) or two objects from the negative pole (neither object was judged to be *big* in the SJT). This is expected: if children know the meaning of *big*, then they know that it denotes the property of having a size that exceeds the standard of comparison in the context. If the context is such that (a) there are only two objects under consideration, and (b) there is a presupposition that only one of them has the property denoted by *big* (thanks to the use of *the*), then assigning the right standard to *big* should be automatic. In contrast, the fact that children systematically took significantly longer to respond to a request for *the full/straight one* in contexts involving two objects that did not satisfy the maximal standard (that is, which were not maximally full or straight) suggests that they were engaged in an interpretive process distinct from what is involved in assigning the correct contextual valuation to a relative GA.

Crucially, if the same kind of reasoning were at play in children's evaluation of requests for *the full one* in the 'not full/less full' context (or *the straight one*), we would have observed similar RTs. That is, if *full* meant 'fuller than the contextual standard' or 'at least as close to maximally full as the contextual standard', then the choice between two objects of unequal fullness would also be straightforward: if context dictates that just one of the objects should satisfy the property, it will always be the fuller of the two. The fact that children took significantly longer to respond to such requests, even though here, too, they targeted the object with the greater degree of the relevant property, shows that the same kind of reasoning is not at work in evaluating absolute GAs such as *full* in such contexts, and therefore that the kind of interpretive variability that such adjectives display – while *sensitive* to the context – is not indicative of a *context-dependent* semantics.

But what kind of reasoning do the long RTs indicate? Our hypothesis is that children took longer to respond to such cases precisely because they were aware that neither of the

objects in question actually satisfied the description used in the request, and that this knowledge triggered an evaluation of whether one of the objects was actually close enough to having the property in order for them to treat the speaker's description as though it were a description of that object. There are different ways to formally characterize this notion of imprecision. For concreteness, we will assume with Lasersohn (1999) that it involves (a) a computation of a set of alternative denotations for an expression that are ordered relative to the actual meaning, and (b) a decision about how far down this list one can go before getting too far away from the actual denotation. In contexts in which the actual meaning of an utterance fails to apply, if there is an element among the set of 'tolerable alternatives' that could be true, then the utterance can be taken to count as 'close enough to true' for the purpose of the discourse, and treated as though it had been an utterance of the tolerable alternative. Crucially, this kind of reasoning is fundamentally *pragmatic* in nature, as it involves judgments about communicative intent and whether a particular utterance can be used in a way that (strictly speaking) conflicts with the truth conditional requirements that it imposes by virtue of its semantics. It is this move beyond the computation of semantic content – even content that invokes contextual information – that, we claim, imposes an extra processing load and results in the longer RTs we observed.

If this is correct, then we have experimental evidence for a distinction between two types of interpretive variability. One type – that exhibited by relative GAs in the positive form – is fundamentally *semantic* in nature, and is based on the conventional meaning of particular expressions (or combinations thereof). A second type – that exhibited by imprecise uses of maximum standard absolute GAs – is fundamentally *pragmatic*, and involves computation of a set of alternative meanings and a contextual judgment about how many of them count as tolerable deviations from the actual, precise meaning of the expression. The differences between children and adults that we observed in their willingness to accept false

descriptions based on maximum standard GAs *full* and *straight* can be explained by assuming that children are more willing to tolerate imprecision than adults, at least in this experimental task.[20] If our overall interpretation of the data is correct, then we predict that smaller deviations from the maximum standard should result in adults patterning as the children did – accepting false descriptions based on absolute maximum standard GAs such as *full* and *straight*, but taking longer to do so than when accepting a description based on a relative GA.

7. Conclusion

In this paper, we have explored how the appearance of gradable adjectives (GAs) in definite descriptions sheds light on the various ways in which context and meaning interact and what young children and adults know about such interaction. Our findings therefore address open questions in the field of language acquisition and development concerning children's understanding of fine-grained distinctions within the adjectival category and their lexical representation of the definite determiner, at the same time as they provide psycholinguistic evidence for theoretical claims made in the field of semantics concerning these representations.

---

[20] One potential objection to this interpretation of the results is that preschoolers are generally not imprecise, and can in fact be quite literal at times. (We thank Anastasia Giannakidou, p.c., for raising this point.) Why, then, would they tolerate imprecision in the PAT? In fact, we are not claiming that children are imprecise; indeed, the slower RTs for the non-maximal pairs indicate that their willingness to give the puppet the fuller or straighter member of the pair is far from automatic. Our claim is rather that their desire to maintain a high standard of precision was, in this case, overridden by an even stronger desire to respond to the speaker's request and find a way to allow for it to be felicitous in the context. Evidently, the opposite is true for adults, though as we mentioned above, we suspect that the results would change if the deviation from the maximal standard were reduced.

Our results open a host of interesting and important questions for future research in language learning. For example, what kinds of evidence lead learners to classify an adjective as gradable, and subsequently as relative or absolute? (See Syrett (2007) for an account of how learners may recruit probabilistic information in the exposure language to classify novel adjectives into GA subclasses.) How does a learner decide whether context is incorporated into the semantics of an adjective or whether it enters as part of the pragmatics? To what extent do these distinctions follow directly from the mapping of an adjective to an appropriate concept as opposed to being linguistically conditioned?

At the same time, these results also provide an interesting methodological lesson. While the debate concerning the ways in which context and meaning interact has raged for years within the fields of linguistics and philosophy, resisting conclusive evidence, the current experiments allow us to ask about the real-time consequences of alternative approaches. Psycholinguistic evidence may allow us to see how alternative computations connected to different semantic representations unfold in real time. Our work demonstrates that the process of verifying the meaning of an expression can in turn provide evidence for alternative hypotheses about the meaning itself. Moreover, such evidence can be provided by children as well as from adults.

Authors' addresses

Kristen Syrett
Rutgers Center for Cognitive Science (RuCCS)
Psychology Building Addition, Busch Campus
Rutgers University – New Brunswick
152 Frelinghuysen Rd.
Piscataway, NJ 08854-8020

Christopher Kennedy
Department of Linguistics
University of Chicago
1010 E. 59th St.
Chicago, IL 60637

Jeffrey Lidz
Department of Linguistics
1401 Marie Mount Hall
University of Maryland
College Park, MD 20742-7505

References

Abbott, B. (1999), Support for a unique theory of definite descriptions. In T. Matthews and D. Strolovitch (eds.), *Proceedings from Semantics and Linguistic Theory IX*. CLC Publications. Ithaca, NY. 1-15.

Barner, D. & J. Snedeker (2007), Compositionality and statistics in adjective acquisition: 4-year-olds interpret *tall* and *short* according to the typical sizes of novel noun referents. In H. Caunt-Nulton, S. Kulatilake, and I. Woo (eds.), *Proceedings of the 31st Annual Boston University Conference on Language Development*. Cascadilla Press. Somerville, MA. 81-92.

Barner, D. & J. Snedeker (2008), Compositionality and statistics in adjective acquisition: 4-year-olds interpret tall and short based on the size distributions of novel noun referents. *Child Development* **79**: 594-608.

Bartlett, E. (1976), Sizing things up: The acquisition of the meaning of dimensional adjectives. *Journal of Child Language* **3**: 205-219.

Bartsch, R. & T. Vennemann (1972), The grammar of relative adjectives and comparison. *Linguistische Berichte* **20**: 19-32.

Birner, B. & G. Ward (1994), Uniqueness, familiarity, and the definite article in English. In S. Gahl, A. Dolbey, and C. Johnson (eds.), *Proceedings of the 20th Annual Meeting of the Berkeley Linguistics Society*. Berkeley Linguistics Society. Berkeley, CA. 93-102.

Booth, A. E. & S. R. Waxman (2003), Mapping words to the world in infancy: Infants' expectations for count nouns and adjectives. *Journal of Cognition and Development* **4**: 357-381.

Brewer, W. & J. B. Stone (1975), Acquisition of spatial antonym pairs. *Journal of Experimental Child Psychology* **19**: 299-307.

Carey, S. (1978), The child as word learner. In M. Halle, J. Bresnan, and G. Miller (eds.),

Linguistic Theory and Psychological Reality. MIT Press. Cambridge, MA. 264-293.

Clark, E. V. (1972), On the child's acquisition of antonyms in two semantic fields. *Journal of Verbal Learning and Verbal Behavior* **11**: 750-758.

Clark, E. V. (1973), What's in a word? On the child's acquisition of semantics in his first language. In T. Moore (ed.), *Cognitive Development and the Development of Language* Academic Press. New York, NY. 65-110. Reprinted (1977) in C. Snow and B. Ferguson (eds.), *Talking to Children*. Cambridge University Press. Cambridge, MA. 186-233.

Clark, H. (1973), Space, time, semantics, and the child. In T. Moore (ed.), *Cognitive Development and the Acquisition of Language.* Academic Press. New York, NY. 27-63.

Coley, J. D. & S. A. Gelman (1989), The effects of object orientation and object type on children's interpretation of the word "big". *Child Development* **60**: 372-380.

Cresswell, M. J. (1976), The semantics of degree. In B. Partee (ed.), *Montague grammar*. Academic Press. New York. 261-292.

Cruse, D. *Lexical semantics.* Cambridge University Press. Cambridge.

Diesendruck, G., D. G. Hall, & S. Graham (2006), Children's use of syntactic and pragmatic knowledge in the interpretation of novel adjectives. *Child Development* **77**: 16-30.

Donaldson, M. & R. J. Wales (1970), On the acquisition of some relational terms. In J. R. Hayes and R. Brown (eds.), *Cognition and the Development of Language*. John Wiley and Sons, Inc. New York, NY. 235-268.

Ebeling, K. S. & S. A. Gelman (1988), Coordination of size standards by young children. *Child Development* **59**: 888-896.

Ebeling, K. S. & S. A. Gelman (1994), Children's use of context in interpreting "big" and "little". *Child Development* **65**: 1178-1192.

Ehri, L. (1976), Comprehension and production of adjectives and seriation. *Journal of Child Language* **3**: 369-384.

Eilers, R. E., D. Kimbrough Oller, & J. Ellington (1974), The acquisition of word-meaning for dimensional adjectives: The long and short of it. *Journal of Child Language* **1**: 195-204.

Fara, D. G. (2000), Shifting sands: An interest-relative theory of vagueness. *Philosophical Topics* **28**: 45-81. Originally published under the name "Delia Graff".

Finch-Williams, A. (1981), *Biggerest or Biggester: A Study in Children's Acquisition of Linguistic Features of Comparative Adjectives and Nonlinguistic Knowledge of Seriation*. Unpublished Ph.D. thesis. Dept. of Speech-Language-Hearing: Sciences and Disorders. University of Kansas. Lawrence, KS.

Gathercole, V. C. (1979), *Birdies Like Birdseed the Bester than Buns: A Study of Relational Comparatives and Their Acquisition*. Unpublished Ph.D. thesis. Dept. of Linguistics. University of Kansas. Lawrence, KS.

Gelman, S. A. & K. S. Ebeling (1989), Children's use of nonegocentric standards in judgments of functional size. *Child Development* **60**: 920-932.

Gelman, S. & E. M. Markman (1985), Implicit contrast in adjectives vs. nouns: Implications for word-learning in preschoolers. *Journal of Child Language* **6**: 125-143.

Gentner, D. (1982), Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. Kučzaj II (ed.), *Language Development Vol. 2: Language, Thought and Culture*. Lawrence Erlbaum. Hillsdale, NJ. 301-334.

Gitterman, D. & J. R. Johnston (1983), Talking about comparisons: A study of young children's comparative adjective usage. *Journal of Child Language* **10**: 605-621.

Graham, S., A. Welder, & A. McCrimmon (2003), Hot dogs and zavy cats: Preschoolers' and adults' expectations about familiar and novel adjectives. *Brain and Language* **84**: 16-

37.

Graham, S., C. Cameron, & A. Welder (2005), Preschoolers' extension of familiar adjectives. *Journal of Experimental Child Psychology* **91**: 205-226.

Graziano-King, J. (1999), *Acquisition of comparative forms in English*. Unpublished Ph.D. thesis. Graduate Program in Linguistics. City University of New York. New York, NY.

Graziano-King, J. & H. S. Cairns (2005), Acquisition of English comparative adjectives. *Journal of Child Language* **2**: 345-373.

Hall, D. G., S. R.Waxman, & W. Hurwitz (1993), How two- and four-year-old children interpret adjectives and count nouns. *Child Development* **64**: 1651-1664.

Harris, P. L., J. E. Morris, & M. Meerum Terwogt (1986), The early acquisition of spatial adjectives: A cross-linguistic study. *Journal of Child Language* **13**: 335-352.

Heim, I. (1990), E-type pronouns and donkey anaphora. *Linguistics and Philosophy* **13**: 137-177.

Hollich, G. (2003), SuperCoder [Computer program]. Retrieved from http://hincapie.psych.purdue.edu/Splitscreen/SuperCoder.dmg.

Hurewitz, F., S. Brown-Schmidt, K. Thorpe, L. R. Gleitman, & J. C. Trueswell (2000), One frog, two frog, red frog, blue frog: Factors affecting children's syntactic choices in production and comprehension. *Journal of Psycholinguistic Research* **29**: 597-626.

Kadmon, N. (1990), Uniqueness. *Linguistics and Philosophy* **13**: 273-323.

Kamp, H. (1980), A semantics for positive and comparative adjectives. *Linguistics and Philosophy* **4**: 1-45.

Kamp, H. & B. H. Partee (1995), Prototype theory and compositionality. *Cognition* **57**: 129-191.

Karmiloff-Smith, A. (1979), *A Functional Approach to Child Language: A Study of*

Determiners and Reference. *Cambridge Studies in Linguistics Vol. 24*. Cambridge University Press. New York, NY.

Keil, F. & J. Carroll (1980), The child's acquisition of "tall": Implications for an alternative view of semantic development. *Papers and Reports on Child Language Development* **19**: 21-28.

Kennedy, C. (1999), *Projecting the adjective: The syntax and semantics of gradability and comparison*. Garland. New York, NY. (1997 Doctoral dissertation, University of California, Santa Cruz).

Kennedy, C. (2007), Vagueness and grammar: The semantics of relative and absolute gradable adjectives. *Linguistics and Philosophy* **30**: 1-45.

Kennedy, C. & L. McNally (2005), Scale structure, degree modification, and the semantics of gradable adjectives. *Language* **81**: 345-381.

Klatzky, R. L., E. V. Clark, & M. Macken. (1973), Asymmetries in the acquisition of polar adjectives: Linguistic or conceptual? *Journal of Experimental Child Psychology* **16**: 32-46.

Klein, E. (1980), A semantics for positive and comparative adjectives. *Linguistics and Philosophy* **4**: 1-45.

Klein, E. (1991), Comparatives. In A. von Stechow and D. Wunderlich (eds.), *Semantik: Ein Internationales Handbuch der Zeitgen Ösischen Forschung* Walter de Gruyter. Berlin. 673-691.

Klibanoff, R. S. & S. R. Waxman (1998), Preschoolers' acquisition of novel adjectives and the role of basic-level kind. In A. Greenhill, M. Hughes, H. Littlefield and H. Walsh (eds.), *Proceedings of the 22nd Annual Boston University Conference on Language Development.* Cascadilla Press. Somerville, MA. 442-453.

Klibanoff, R. S. & S. R. Waxman (2000), Basic level object categories support the

acquisition of novel adjectives: Evidence from preschool-aged children. *Child Development* **71**: 649-659.

Kyburg, A. and M. Morreau (2000), Fitting words: Vague language in context. *Linguistics and Philosophy* **23**:577-597.

Lasersohn, P. (1999), Pragmatics halos. *Language* **75**: 522-551.

Layton, T. L. & S. L. Stick (1978), Comprehension and production of comparatives and superlatives. *Journal of Child Language* **6**: 511-527.

Lewis, D. (1968), Counterpart theory and quantified modal logic. *Journal of Philosophy* **65**: 113-126.

Lewis, D. (1979), Scorekeeping in a language game, *Journal of Philosophical Logic* **8**: 339-359.

Ludlow, P. (1989), Implicit comparison classes. *Linguistics and Philosophy* **12**: 519-533.

Macnamara, J. (1972), Cognitive basis of language learning in infants. *Psychological Review* **79**: 1-13.

Maratsos, M. (1976), *The Use of Definite and Indefinite Reference in Young Children: An Experimental Study of Semantic Acquisition*. Cambridge University Press. New York, NY.

Marschark, M. (1977), Lexical marking and the acquisition of relational size concepts. *Child Development* **48**: 1049-1051.

Meroni, L. (2006), *Putting Children in Context: Experimental Studies on Children's Interpretation of Definite Noun Phrases*. Unpublished Ph.D. thesis. Dept. of Linguistics. University of Maryland. College Park, MD.

Mintz, T. (2005), Linguistic and conceptual influences on adjective acquisition in 24- and 36-month olds. *Developmental Psychology* **41**: 17-29.

Mintz, T. & L. R. Gleitman (2002), Adjectives really do modify nouns: The incremental and

restricted nature of early adjective acquisition. *Cognition* **84**: 267-293.

Moore, D. (1999), Comparatives and superlatives: Lexical before functional. In A. Greenhill, H. Littlefield, and C. Tano (eds.), *Proceedings of the 23rd Annual Boston University Conference on Language Development.* Cascadilla Press. Somerville, MA. 474-481.

Nadig, A., J. Sedivy, A. Joshi, and H. Bortfeld (2003), The development of discourse constraints on the interpretation of adjectives. In B. Beachley, A. Brown, and F. Conlin (eds.), *Proceedings of the 27th Annual Boston University Conference on Language Development*. Cascadilla Press. Somerville, MA. 568-579.

Neale, S. (1990), *Descriptions.* MIT Press. Cambridge, MA.

Nelson, K. (1976), Some attributes of adjectives used by young children. *Cognition* **4**: 1-31.

Nelson, K. & H. Benedict (1974), The comprehension of relative, absolute, and contrastive adjectives by young children. *Journal of Psycholinguistic Research* **3**: 333-342.

Prasada, S. (1992), *Acquisition of adjective meanings: A lexical semantic approach*. Unpublished Ph.D. Thesis. Dept. of Brain and Cognitive Sciences. MIT. Cambridge, MA.

Prasada, S. & M. Cummins (2001), Structural constraints on the interpretation of novel count nouns. In A. H-J. Do, L. Dominguez, and A. Johansen (eds.), *Proceedings of the 25th Annual Boston University Conference on Language Development*. Cascadilla Press. Somerville, MA. 623-632.

Raffman, D. (1994), Vagueness without paradox. *The Philosophical Review* **103**: 41-74.

Raffman, D. (1996), Vagueness and context relativity. *Philosophical Studies* **81**: 175-192.

Ravn, K. E. & S. Gelman (1984), Rule usage in children's understanding of "big" and "little". *Child Development* **55**: 2141-2150.

Rips, L. & W. Turnbull. (1980), How big is big? Relative and absolute properties in memory. *Cognition* **8**: 145-174.

Roberts, C. (2003), Uniqueness in definite noun phrases. *Linguistics and Philosophy* **26**: 287-350.

Rotstein, C. & Y. Winter (2004), Total adjectives vs. partial adjectives: Scale structure and higher-order modifiers. *Natural Language Semantics* **12**: 259-288.

Rusiecki, J. (1985), *Adjectives and comparison in English*. Longman linguistics library, Vol. 31. Longman Group Limited. London.

Russell, B. (1905), On denoting. *Mind* **14**: 479-493.

Sena, R. & L. B. Smith (1990), New evidence on the development of the word "big". *Child Development* **61**: 1034-1052.

Sera, M. & L. B. Smith (1987), "Big" and "little": "Nominal" and relative uses. *Cognitive Development* **2**: 89-111.

Sharpe, D., I. Fonte, & E. Christe (1998), Big mice, big animals, big problems: The acquisition of adjective interpretation rules. In A. Greenhill, M. Hughes, H. Littlefield, and H. Walsh (eds.), *Proceedings of the 22nd Annual Boston University Conference on Language Development*. Cascadilla Press. Somerville, MA. 675-683.

Smith, L. B., N. Cooney, & C. McCord (1986), What is "high"? The development of reference points for "high" and "low". *Child Development* **57**: 583-602.

Soames, S. (1999), *Understanding Truth*. Oxford University Press. Oxford.

Sperber, D. & D. Wilson (1986), Loose talk. *Proceedings of the Aristotelian Society New Series* **86**: 153-171.

Strawson, P. F. (1950), On referring. *Mind* **59**: 320-344.

Syrett, K. (2007). *Learning about the Structure of Scales: Adverbial Modification and the Acquisition of the Semantics of Gradable Adjectives*. Unpublished Ph.D. thesis. Dept. of Linguistics. Northwestern University. Evanston, IL.

Syrett, K., J. Lidz, & C. Kennedy (2006a), The semantic typology of gradable adjectives:

Experimental evidence from adult and child language. Poster presented at the *CUNY Human Sentence Processing Conference*. March 23-25.

Syrett, K., J. Lidz, & C. Kennedy (2006b), Scalar structure and the semantic representations of gradable adjectives in child language. Paper presented at the *Main Session of the 42nd Annual Meeting of the Chicago Linguistic Society.* April 6-8.

Syrett, K., E. Bradley, J. Lidz, & C. Kennedy (2006), Shifting standards: Children's understanding of gradable adjectives. In K. Ud Deen, J. Nomura, B. Schulz, and B. D. Schwartz (eds.), *Proceedings of the Inaugural Conference on Generative Approaches to Language Acquisition - North America*. Cambridge, MA: UConn Occasional Papers in Linguistics Vol. 4. Honolulu, HI. 353-364.

Taylor, M. & S. Gelman (1988), Adjectives and nouns: Children's strategies for learning new words. *Child Development* **59**: 411-419.

Townsend, D. (1976), Do children interpret "marked" comparative adjectives as their opposites? *Journal of Child Language* **3**: 385-396.

Trueswell, J. C., I. Sekerina, N. M. Hill, & M. L. Logrip (1999), The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition* **73**: 89-134.

Unger, P. (1975), *Ignorance*. Clarendon Press. Oxford.

von Stechow, A. (1984), Comparing semantic theories of comparison. *Journal of Semantics* **3**: 1-77.

Waxman, S. R. & A. E. Booth (2001), Seeing pink elephants: Fourteen-month-olds' interpretations of novel nouns and adjectives. *Cognitive Psychology* **43**: 217-242.

Waxman, S. R. & R. S. Klibanoff (2000), The role of comparison in the extension of novel adjectives. *Developmental Psychology* **36**: 571-581.

Waxman, S. R. & D. Markow (1995), Words as invitations to form categories: Evidence from 12- to 13-month-old infants. *Cognitive Psychology* **29**: 257-302.

Waxman, S. R. & D. Markow (1998), Object properties and object kinds: Twenty-one-month-old infants' extension of novel adjectives. *Child Development* **69**: 1313-1329.

Wexler, K. (2003), Maximal trouble. Paper presented at the *CUNY Conference on Human Sentence Processing*. March 27-29.

Wheeler, S. (1972), Attributives and their modifiers. *Noûs* **6**: 310-334.