Meaning and Context in Children’s Understanding of Gradable Adjectives

Kristen Syrett\textsuperscript{a}, Christopher Kennedy\textsuperscript{b}, Jeffrey Lidz\textsuperscript{c}

\textsuperscript{a} Rutgers University, \textsuperscript{b} University of Chicago, \textsuperscript{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 meanings – relative, absolute maximum standard, and absolute minimum standard – based on the role of the context in setting the standard of comparison. These results lend support to a typology of gradable adjectives determined by abstract differences in scalar structure. While relative gradable adjectives such as *big* depend on the context for the standard of comparison, absolute gradable adjectives (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 definite descriptions and measure their willingness to accommodate when these presuppositions are violated. An analysis of children’s reaction times in their non-adult-like acceptance of a puppet’s infelicitous requests with maximal standard GAs (*full*, *straight*) reveals an allowance of imprecision in the degree to which the standard must hold that we argue is pragmatic in nature, and therefore distinct from the vagueness encoded in the semantic representations of relative GAs.
Meaning and Context in Gradable Adjectives

1. Introduction

A significant part of becoming a competent language user involves understanding the relation between context and meaning. In this paper, we explore three ways that meaning and context interact in child and adult language: the interpretation of expressions whose meanings are partially determined by the context of utterance (semantic context dependence); context-based conditions on the felicitous use of an expression (presupposition accommodation); and informative uses of expressions in contexts in which they strictly speaking do not apply (imprecision). Our empirical focus involves uses of gradable adjectives in definite descriptions (noun phrases), in which each of these three types of context/meaning interaction is manifested. Specifically, we focus on definite descriptions in which a gradable adjective modifier takes on different interpretations in different contexts in order to ensure that the presuppositions of existence and uniqueness introduced by the definite determiner are satisfied. We show that by the age of three, children are sensitive to all three varieties of context dependence mentioned above and that their application of the different context/meaning relations is appropriately guided by the relevant aspects of their linguistic representations. Thus we demonstrate that by age three, children’s semantic representations for gradable adjectives and definite determiners are fully adult-like and moreover, that the consequences of these representations for relating meaning and context are in place.

1.1 Relative Gradable Adjectives and Context Dependence

Gradable adjectives (GAs) are adjectives whose core meaning involves reference to a scalar concept on the basis of which objects can be ordered (e.g., height, weight, cost, etc.). Distributionally, gradable adjectives 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 gradable adjective is located on the corresponding scale. Our focus in this paper is the meaning and use of gradable adjectives in contexts that lack degree morphology of any kind: GAs in the so-called positive form, such as tall, heavy and expensive (and not their explicit comparative form). From a semantic perspective, the positive form is interesting because it is context dependent: intuitively, what ‘counts as’ tall can vary from context to context, depending on the syntactic environment in which it appears, the topic of discussion, or the interests and expectations of the participants in the discourse.

For example, although there is clearly a shared meaning between the three uses of tall in (1a-c) (which we will characterize in more detail below), it is equally clear that on the most natural interpretations of these examples, the actual height that is sufficient to qualify as tall changes: tall snowmen are shorter than tall buildings, and tall buildings are shorter than tall mountains.

(1)   a. That is a tall snowman.
     b. That is a tall building.
     c. That is a tall mountain.

The examples in (2a-b) involve similar variation in what it takes to count as tall, but in this case it is not the modified noun that provides the basis for this shift, but rather extralinguistic knowledge about the heights of snowmen that is brought to bear in light of the meanings of the agents of the described events: third graders tend to build smaller snowmen than fraternity brothers (Kamp & Partee, 1995).

(2)   a. The third graders built a tall snowman.
     b. The fraternity brothers built a tall snowman.

Finally, we see context dependence at work even in a simple predication such as (3a):

we might know that Anna is a woman, a gymnast, and taller than the average gymnast, but whether (3a) is judged to be true or false of Anna depends on how we understand the
adjective. If we understand (3a) as spelled out in (3b), then we would probably judge the sentence to be true. If we understand (3a) as in (3c), however, we might very well judge it to be, since even a tall gymnast might be a short woman.

(3)  
   a. Anna is tall.  
   b. Anna is tall for a gymnast.  
   c. Anna is tall for a woman.  

In what follows, we will refer to GAs that are context dependent in the positive form as \textit{relative} GAs (to be distinguished later in section 1.3 from a second class of GAs which are 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 possesses a degree of the scalar concept encoded by the GA that exceeds a contextually determined \textit{standard of comparison} (see e.g., Bartsch & Vennemann, 1972; Cresswell, 1977; Kennedy, 1999, 2007; Klein, 1980, 1991; Ludlow, 1989; von Stechow, 1984; Wheeler, 1972; and many others). The standard of comparison is a degree that typically corresponds to an ‘average’ or ‘norm’ of the measured concept relative to some salient set (usually referred to as a ‘comparison class’), which may be explicitly indicated by a constituent that the adjective is in construction with (as in (1a-c) and (3b-c)), may be inferred based on other information in the sentence (as in (2a-b)), or may be implicit (as in (3a)). (The standard of comparison can also be fixed on a more arbitrary basis in special contexts; see Fara (2000) for discussion.)

There are a number of ways in which this kind of analysis can be implemented in a compositional semantics, which differ primarily on the semantic type assigned to lexical adjectives and in more general assumptions about the relation between the morphologically unmarked positive form and the various forms that have explicit degree morphology. For the purposes of this paper, we will assume a simplified semantics that abstracts away from these issues, showing just those features of the positive form that a more fully articulated
compositional account will have to end up with. Specifically, we will assume that relative GAs in the positive form have context-dependent denotations along the lines of the one given for tall in (4), where \([[\alpha]]^c\) means ‘the denotation of \(\alpha\) in context \(c\)’, and \(\lambda x. \phi\) means ‘the property that is true of an object \(o\) just in case \(\phi\) is true when \(o\) is substituted for all occurrences of \(x\) in \(\phi\).’

\[
(4) \quad [[\text{tall}]]^c = \lambda x. x’s \text{ height exceeds the standard of comparison for height in } c
\]

We will have nothing to say here about how the standard of comparison is actually computed, or differences that may exist between the various kinds of examples illustrated above based on the extent to which the comparison class is made explicit. (See Kennedy (2007) for detailed discussion.) While these are all important issues, they are orthogonal to the point that is of central concern to us here, and which is made explicit in (4): the meaning of a relative GA such as tall varies from context to context, depending on where the standard of comparison is actually set. Specifically, the feature of meaning of tall that changes is its extension: the things of which it is true and false.

Returning to the example in (3a), let us use \(c(\text{gymnast})\) to represent a context in which we are talking about (female) gymnasts, and intend (3a) to be understood as (3b), and \(c(\text{woman})\) to represent a context in which we are talking about women and intend (3a) to be understood as (3c). Whenever the standards of comparison in \(c(\text{gymnast})\) and \(c(\text{woman})\) are distinct (which is normally the case), we will end up with the result that \([[\text{tall}]]^{c(\text{gymnast})} \neq [[\text{tall}]]^{c(\text{woman})}\), because the set of things whose heights exceed the standard in the former context is a proper subset of the set of things whose heights exceed the standard in the latter context. In particular, if Anna’s height falls in between the standard of comparison in \(c(\text{gymnast})\) and \(c(\text{woman})\), as represented graphically in (5), then \([[\text{tall}]]^{c(\text{gymnast})}\) is true of Anna but \([[\text{tall}]]^{c(\text{woman})}\) is false of Anna.

\[
(5) \quad \text{HEIGHT: } 0 \quad \text{----------- standard}_{c(\text{gymnast})} \quad \text{----- height}_{\text{Anna}} \quad \text{----- standard}_{c(\text{woman})} \quad \text{-------->}
\]
The denotation in (4) (and comparable denotations for other relative GAs) thus captures the fact that all uses of the positive form involve a shared meaning: a relation between the degree to which an object manifests some scalar concept (in this case, tallness) and a contextually determined standard of comparison for that concept. However, it is not until we actually fix the standard of comparison that that we get a property that can be true or false of an object – the kind of thing that can be used to make assertions or denials, ask questions, and so forth. It is in this sense that the meaning of the positive form is context dependent.

Before moving on, we should explicitly mention that in addition to being context dependent, the positive form is also vague, in the sense that there are contexts in which even with complete knowledge of the relevant facts, it may not be possible to judge the truth of a predication involving it. For example, in the case of (3a), we might know exactly what Anna’s height is, that the intended meaning of (3a) is (3b), and that the average height of a gymnast is such-and-such, and yet still be unsure about whether it is actually true that Anna is tall (for a gymnast): she could be a borderline case. For example, if we know that the average height of a gymnast is 1.45 meters, and Anna is 1.6 meters, then (3b) is clearly true. However, if Anna is 1.5 meters tall, then (3b) does not necessarily follow.

Similarly, we find it difficult or impossible to judge adjacent objects along a scalar continuum differently relative to a positive form GA whose meaning makes reference to that continuum: the conditional in (6), for example, seems to be straightforwardly valid.

\[(6) \quad \text{For any individuals } a \text{ and } b, \text{ if } a \text{ is tall and } b \text{ is just a bit shorter than } a, \text{ then } b \text{ is also tall.} \]

But if (6) were valid, it would support (through successive applications) the unacceptable conclusion that a 1-meter-tall man is tall given the initial (uncontroversial) premise that a 2-
meter-tall man is tall; this is an example of the so-called Sorites Paradox (the paradox of the heap).

Vagueness and context dependence go hand in hand in the case of the positive form (of a relative GA), and a full account of the semantic properties of this construction must take both into account. Simply put, the mere assumption that the positive form is context dependent, even in the way outlined above, is not enough to account for vagueness, since denotations along the lines of (4) do not by themselves explain borderline cases or our (mistaken) intuitions about the inductive premise of a Sorites argument. However, most accounts of vagueness make crucial use of context dependence (e.g., Fara, 2000; Kamp, 1981; Kamp & Partee, 1995; Kennedy, 2007; Raffman 1994, 1996; Soames, 1999), and we will assume that ultimately these two aspects of meaning are related features of the semantics of the positive form. To keep the exposition as clear as possible, however, we will continue to talk strictly in terms of the context dependence of the positive form rather than its vagueness, both because the crucial factor that is important for our experimental investigation of the context/meaning relation is the way that different contexts result in different standards of comparison, and because the other features of vagueness outlined above do not play a role in our investigation.

1.2 Gradable Adjectives in Definite Descriptions: Accommodating the Standard of Comparison

The context dependence of relative GAs provides a great deal of flexibility in their use: because they can convey different properties in different contexts, unified by their ‘relation to a standard’ semantic core, the same form can be used to convey different information about the degree to which an object manifests some scalar concept, as illustrated by the examples in (1)-(3) above. A particularly striking example of this flexibility comes
from the use of relative GAs in definite descriptions, which is the empirical focus of this paper.

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 NP of the form the φ is associated with 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 & 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 if she accepts such a use.

The effect of these presuppositions on judgments of felicity can be illustrated by the following example. Consider a context in which two individuals A and B are sitting across 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 (7) 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 denoted by the nominal constituent blue rod.

(7) #Please give me the blue rod.\(^1\)

\(^1\) In this paper, we will use the ‘#’ symbol to indicate an utterance that is grammatical, but infelicitous.
By the same token, A’s utterance of (8) would be just as infelicitous, in this case because the existence presupposition of the definite is not met: there is no object that satisfies the property red rod in the context.

(8) #Please give me the red rod.

Speaker A can, however, felicitously use (9) to request the longer of the two rods.

(9) Please give me the long rod.

Importantly, (9) is felicitous when the two rods are independently judged to be long, when they are independently judged to be not long, and when one is considered long and the other not: all that matters is that there is a difference in length between them. On the surface, this appears to conflict with what we saw with (7), which fails uniqueness because both rods are blue rods, and (8), which fails existence because neither rod is a red rod. The crucial difference, of course, is that long is a positive form relative GA, and so denotes the context dependent property in (10).

(10) \[[\text{long}_{\text{pos}}]\]^c = \lambda x. x’s length exceeds the standard of comparison for length in c

Two features are particularly salient about the context in which (9) is used, and these bear on the choice of a standard of comparison: the presuppositions of the definite NP require there to be a unique long rod in the context, and the two rods are of unequal lengths. Given that there are two salient rods in the context c(rod), the only way to satisfy these presuppositions is for one of them to count as long and the other one not, relative to the context. Because the rods have unequal lengths, this result can be achieved by ‘shifting’ the prevailing standard of comparison – the one on the basis of which the rods are independently judged to be both long or both not long – so that it differentiates between them, as represented in (11).

(11) LENGTH: 0 ------------ length_{rod_1} ----- standard_c ----- length_{rod_2} ------------>
This move from the prevailing standard (of what is considered to be a long rod) to one that allows the presuppositions of the definite to be satisfied in that context is an instance of presupposition accommodation (Lewis, 1968) – treating the context as though it contained the information relevant to satisfying some presupposition. In this case, the crucial move is to treat the context as though the standard of comparison for long is one that makes it true of one rod and false of the other. Given the statement of the truth conditions of the positive form GAs, this amounts to making it true of the longer of the two rods (if one object has a degree of the scalar concept named by a gradable adjective $\alpha$ that exceeds the standard for $\alpha$, and a second object does not, then the first is more $\alpha$ than the second).\footnote{There is a way of using a positive form GA in a definite description to ask for the lower-ranked object: use the polar-negative antonym, e.g., the short rod. This strategy does not always work due to markedness effects associated with negative members of antonym pairs, but for those adjectives that do not show such effects (or show them weakly, such as short), the basic account runs the same as for long, modulo the difference between the antonyms in the orderings they impose: one is the inverse of the other, so ‘more shortness’ corresponds to ‘less longness’.
}

In sum, definite descriptions involving nouns modified by positive form relative GAs are particularly interesting for studies of the context/meaning interface because it is possible to construct scenarios (the differentiation tasks described above) in which they invoke two different features of this interface: 1) contextually adjusting the meaning (extension) of the adjective by manipulating the standard of comparison, and 2) doing so in such a way that accommodates the presuppositions of the definite description.

1.3 Absolute Gradable Adjectives and Imprecision

As it turns out, not all GAs are context dependent (or vague) in the positive form, a fact that provides the basis for a set of control data in experimental tasks that use facts like...
those discussed in the previous section to investigate understanding of the context/meaning interface. In addition to the large set of relative GAs, there is a class of adjectives that are demonstrably gradable (in that they readily allow comparisons) but which do not have context-dependent standards of comparison in the positive form. Following Unger (1975), Rusiecki (1985), Kennedy and McNally (2005), and Kennedy (2007), and others, we will refer to these as absolute GAs (see also Rotstein & Winter, 2004).³ Absolute GAs come in two varieties: minimum standard absolute GAs, which require their arguments to merely possess some degree of a gradable property (e.g., spotted and bent), and maximum standard absolute GAs, which require their arguments to possess a maximal degree of a gradable property (e.g., full and straight). Rusiecki (1985), Cruse (1986), Rotstein and Winter (2004), Kennedy and McNally (2005), Kennedy (2007), and Syrett (2007) provide extensive arguments on the basis of modifier selection and entailment patterns for the absolute/relative distinction; what is important for our purposes here is that neither type of absolute GA has the sort of ‘differentiation use’ in definite descriptions that we see with relative GAs.⁴

For example, maintaining the kind of two-object context discussed in the previous section, a definite description such as the spotted disk can be felicitously used to request one of the two disks just in case one has spots and the other does not, but not to pick out the more

³ Rips and Turnbull (1980) also make use of a relative/absolute distinction among adjectives in their experimental work, but mean something slightly different.

⁴ A question that we will not address in this paper concerns the factors that determine whether a GA is relative or (minimum/maximum) absolute, as these are discussed in detail in other work (cf. Kennedy & McNally, 2005; Kennedy, 2007; Rotstein & Winter, 2004; Syrett, 2007). In general, a GA is relative if the scalar concept it encodes is open (has no natural endpoints), and absolute if it is closed (has either a minimum value or a maximum value or both).
spotted of two disks that are both spotted (or, obviously, one of two spotless disks). Likewise, the full jar can be felicitously used to request one of two jars only if one jar is completely full and the other is not; it cannot normally be used to refer to the fuller of two partially-full jars, or one of two jars that are both full. These judgments are expected if the positive forms of spotted and full have denotations along the lines of (12a) and (12b), respectively.

(12) a. \([\text{spotted}_\text{pos}]^c = \lambda x. x \text{ has some spots}\]
    b. \([\text{full}_\text{pos}]^c = \lambda x. x \text{ has maximal fullness}\]

Unlike the kinds of denotations we posited for relative GAs in the positive form, these denotations are not context dependent: spotted denotes the property of having some spots, and full denotes the property of being maximally full. If this is right, then the difference between the contexts described above that result in infelicity with these adjectives, and the corresponding contexts involving relative GAs, is that the former do not license the possibility of accommodation: if full means ‘maximally full’, independent of context, there is no way to ‘recalibrate’ the meaning of the adjective (by manipulating a standard of comparison) to make the adjective true of an object that is not maximally full, but merely fuller than another.

The initial reaction to these comments is typically to reject them. For example, there is a strong intuition that full merely requires that an object be ‘close to full’, and that in different contexts different approximations may do. 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, it is not so clear what to say about one about one that is 7/8 full, or 15/16 full, or any noticeably-not-but-almost-there-amount full. While it is clear that at some point we

---

5 We intend the fractions here to be based on whatever is the conventional maximum or ‘full-line’ for the container in question. This may be the rim, or it may be a particular mark on the
would typically be willing to start calling the jar full (and would therefore be willing to refer to it as the full jar), and that this point might be different in different contexts (based on our goals, the types of jars, the contents of the jars, etc.), what remains unclear is whether it follows from this that full really is context dependent in the same way as tall and long. Or does it indicate a different kind of context/meaning interaction, one that regulates imprecision rather than vagueness – tolerance of the false-but-informative application of a description to an object in contexts in which such applications are ‘close enough to true’?

There are different answers to this question in the literature, with no clear consensus. However, the experiments we report in this paper provide important new data bearing on this debate, both because they clearly support the core semantic distinction between relative and absolute GAs – the former have context sensitive denotations in the positive form and the latter do not – and because they further indicate that imprecise uses of absolute GAs are processed very differently from uses of relative GAs in which a standard of comparison is shifted to accommodate the presuppositions of a definite NP, suggesting that the two uses do not in fact involve the same kind of context/meaning interaction.

1.4 Overview of the Paper

This paper is organized as follows. First, we examine children’s understanding of gradable adjectives in definite descriptions to probe their level of knowledge of the first two types of context/meaning interaction. We begin by providing some background on children’s knowledge of gradable adjectives and the presuppositions of definite NPs, and then move on to describe our experiments on children’s understanding of differentiation uses of definite descriptions with gradable adjective modifiers. We will see that children as young age three have adult-like competence in the interpretation of both relative and absolute GAs, assigning container, or it may be something less concrete. The subsequent comments apply regardless of which option we choose.
the former context-dependent and the latter fixed interpretations, and that they are sensitive to – and can correctly accommodate – the presuppositions of definite NPs.

We then focus on important differences between children and adults involving maximum standard absolute GAs. Given a request with a definite NP of the form the A one, where A is a maximum standard GA, and a context in which neither of the two objects exhibits the relevant property to a maximum degree, adults in our experiments uniformly rejected the request, illustrating their implicit acknowledgement of the failure of the existence presupposition. Children, however, systematically accepted such uses, subject to other discourse conditions. Crucially, though, they also took significantly longer to accept such requests than those with relative GAs or true predications of maximum or minimum standard absolute GAs. We argue that this finding supports a distinction between true semantic context dependence on the one hand, and contextually acceptable degrees of imprecision on the other. The former is a matter of the conventional meaning of the positive form of relative GAs, which happens automatically because of the kind of meaning that these GAs have; the latter involves pragmatic reasoning about how much deviation from the conventional meaning of a term (a maximum standard absolute GA in this case) should be tolerated in contexts in which it does not hold.

Our study is important in two respects. First, it provides new insights on when children have achieved adult-like levels of competence in two fundamental aspects of the context/meaning relation (semantic context dependence and presupposition accommodation) – by three years of age. Second, it provides new data from child language that bears on more general questions about the nature of contextually-influenced interpretive variability in both children and adults. These data support a distinction between variability that is fundamentally semantic in nature (context dependence) and variability that is fundamentally pragmatic (imprecision).
2. Adjectives and Definiteness in Child Language

2.1 Children’s Knowledge of Adjectives and Contextual Variability

Research on the acquisition of adjectives has largely focused on three questions. First, when do children recognize that adjectives license different inferences than nouns? The evidence points to two conclusions: (a) that noun learning precedes adjective learning, and (b) that well before reaching preschool, children recognize that adjectives refer to object properties and not object kinds (or said another way, that adjectives and nouns denote different kinds of properties) (cf. Booth & Waxman, 2003; Gentner, 1982; Macnamara, 1972; Waxman & Booth, 2001; Waxman & Markow, 1995, 1998, and others). Second, to what extent does knowledge of the category to which a novel adjective is applied impact children’s ability to acquire that word? Here, the evidence suggests that familiarity with the object kind makes it easier for children to acquire an adjective referring to a property of that kind (cf. Hall, Waxman, & Hurwitz, 1993; Klibanoff & Waxman, 1998, 2000; Mintz, 2005; Mintz & Gleitman, 2002; Taylor & Gelman, 1988; Waxman & Klibanoff, 2000). Third, to what extent do modificational (prenominal) uses of adjectives license different inferences than predicative uses? In this domain, evidence suggests that adjectives in the prenominal position may have a privileged role in cueing contrast among object properties (Diesendruck, Hall, & Graham, 2006; Nelson, 1976; Prasada, 1992; Prasada & Cummins, 2001; but see Nadig, Sedivy, Joshi, & Bortfeld, 2003). Such results lay an important foundation to our questions, namely, when do children become aware of the different subcategories of GAs (cf. Graham, Welder, & McCrimmon, 2003, Graham, Cameron, & Welder, 2005) and how do these subcategories differ with respect to their context-dependent features?

With respect to context dependence, preschoolers have been shown to use a variety of information sources from the context to evaluate the use of an adjective. For example, they can use the orientation of an object and its intended reference to determine the relevant
dimension for evaluating its size (cf. Coley & Gelman, 1989). Similarly, they allow the interpretation of an adjective to change across kinds (Sharpe, Fonte, & Christe, 1998), and appreciate that information provided by the noun may be relevant to the interpretation of a novel adjective (e.g., Mintz, 2005; Mintz & Gleitman, 2002).

Restricting our attention to relative GAs, two lines of research have demonstrated that preschoolers are able to shift the standard of comparison for relative GAs such as big, tall, high, and low. A series of papers by Gelman and Ebeling have evaluated how 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 (Ebeling & Gelman, 1988, 1994; Gelman & Ebeling, 1989). Smith and her colleagues have probed the factors affecting children’s willingness to relabel an object once the standard has shifted (Sera & Smith, 1987) and their ability to make use of a given range of values for assigning the standard of comparison (Smith, Cooney, & McCord, 1986). Barner & Snedeker (2007) have also shown that four-year-olds are able to shift the standard when additional same-kind objects of different sizes are included in a relevant set. Together, these studies demonstrate that preschoolers take a range of contextual information into account when assigning the standard of comparison for these adjectives. However, since these studies did not include non-relative GAs as controls (see section 1.3 above), it has remained an open question whether children would consistently do this for all adjectives (gradable or not), or would appropriately restrict this behavior to this class of GAs.

The study that has come closest to introducing control adjectives is Nelson and Benedict (1974). Nelson and Benedict asked children aged three to six for their judgments about a series of pictures of objects and found that children exhibited a shorter response latency to the comparative form of adjectives such as tall (e.g., taller) than they did for adjectives such as happy or leafy. The authors interpreted these results as demonstrating that
the former adjective carry an implicit comparison, whereas the others do not. That is, it takes no more time to process the comparative form of *tall* than the positive form, but it takes more time to process *happier* than *happy*. However, because the authors imposed distinctions among the different sets of adjectives that do not directly correspond to semantic distinctions found in natural language, it is difficult to draw conclusions from these results. The current study starts with the distinctions between relative and absolute GAs discussed above and offers behavioral evidence from both children and adults as empirical support for these distinctions.

2.2 Children’s Production and Comprehension of Definite Noun Phrases

Although the presuppositions of definiteness 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 children do not recognize the presuppositions of definiteness (Wexler, 2003), while and their performance in recent eyetracking experiments (cf. Trueswell, Sekerina, Hill, & Logrip, 1999) has suggested that they have difficulty parsing definite NPs. However, while both of these findings suggest that definiteness poses a problem for children, they actually leave unresolved the question of whether children are aware of the presuppositions that a singular definite NP carries.

There is by now a well-known asymmetry in comprehension and production, so 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 representations. Likewise, the results from eye-tracking experiments reveal that children have difficulty using information from the context to restrict the reference of the noun (cf. Hurewitz, Brown-Schmidt, Thorpe, Gleitman, & Trueswell, 2000; Meroni, 2006),
not that they do not understand what presuppositions are involved. By using definite
descriptions to draw out the distinctions between relative and absolute GAs, we demonstrate
not only that children have different representations for these two kinds of GAs (one being
context-dependent and the other not), but that they see the presuppositions of the singular
definite NP as highlighting these differences in the manner outlined in sections 1.2 and 1.3.

3. Pre-Experiment Scalar Judgment Task

3.1 Introduction

The goal of the Scalar Judgment Task (SJT) was twofold. First, we wanted to elicit
scalar judgments from participants in the hopes that these judgments would reflect a division
among the three subclasses of GAs. Second, we sought to assess children’s judgments of big
and long on stimuli that would be used in Experiments 1 and 2.

3.2 Method

Participants

Thirty-six children representing three age groups participated in this task: 12 three-
year-olds (6 boys 6 girls, range: 3;3 to 3;11, M: 3;8); 12 four-year-olds (5 boys 7 girls, range:
4;1 to 4;11, M: 4;5); and 12 five-year-olds (5 boys 7 girls, range: 5;0 to 5;11, M: 5;5). In
addition, 28 adult native speakers of English (Northwestern undergraduates fulfilling an
experimental requirement for a Linguistics course) served as controls.
Meaning and Context in Gradable Adjectives

Materials

The materials consisted of four sets of seven items each, as outlined in Table 1.

Table 1: Stimuli for Scalar Judgment Task

<table>
<thead>
<tr>
<th>adjective</th>
<th>stimuli</th>
</tr>
</thead>
<tbody>
<tr>
<td>relative</td>
<td></td>
</tr>
<tr>
<td>big</td>
<td>7 wooden blocks painted blue, decreasing incrementally in size from $2^{3/4}$ to $5/8^{3/4}$</td>
</tr>
<tr>
<td>long</td>
<td>7 wooden rods painted green, .5″ in both width and height, decreasing incrementally in length from 8″ to 2″</td>
</tr>
<tr>
<td>absolute</td>
<td></td>
</tr>
<tr>
<td>spotted</td>
<td>7 wooden disks painted red, all 3.75″ in diameter and .5″ thick, ranging from being covered with spots to having no spots</td>
</tr>
<tr>
<td>full</td>
<td>7 clear plastic containers with white lids, all 2″ in height and 1.5″ in diameter, and ranging from being full (of lentils) to empty</td>
</tr>
</tbody>
</table>

Procedure

Participants were presented with four sets of seven objects one after the other. For each item in each set, the experimenter asked the participant, *Is this A?*, where A was a target adjective corresponding to the property exemplified by the objects in the set (*big*, *long*, *full*, *spotted*). The experimenter always started at the positive, rather than the negative, pole (i.e., at the big end, not the small end).\(^6\) There were two conditions, based on the order of

\(^6\) 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 & Snedeker, 2007; Bartlett, 1976; Brewer & Stone, 1975; Carey, 1978; Clark, 1972; Clark, 1973; Eilers,
presentation of the sets. Half of the participants were randomly assigned to the ‘Relative-Absolute’ condition and saw the sets in the following order: big, long, full, spotted (i.e., relative GA, relative GA, absolute GA, absolute GA). The other half of the participants were randomly assigned to the ‘Absolute-Relative’ condition and saw the sets in the following order: full, spotted, big, long. The task took approximately 10 minutes.

3.3 Results

The judgments from adults and children for the four scales are captured in Figure 1 and Figure 2, respectively. The seven items are indicated on the x-axis (greatest degree = 1, least degree = 7), and the percentage of acceptance is on the y-axis.

Kimbrough Oller, & Ellington, 1974; Keil & Carroll, 1980; Klatzky, Clark, & Macken, 1973; Marschark, 1977; Townsend, 1976). In addition, we do not address the apparent decrement in children’s understanding of the term big or the weight of semantic or dimensional features in evaluating the size of an object (cf. Bausano & Jeffrey, 1975; Clark, 1973; Coley & Gelman, 1989; Gathercole, 1982; Harris & Folch, 1985; Harris, Morris, & Meerum Terwogt, 1986; Lumsden & Poteat, 1968; Maratsos, 1973, 1974; Ravn & Gelman, 1984; Sena & Smith, 1990).
For both age groups, there was a decrease in the acceptance for both relative GAs around the midpoint of the series. Judgments for spotted were clear-cut for both children and adults: both children and adults accepted as spotted any disk with any number of spots on it. Judgments for the maximum standard absolute GA full were noticeably different from the minimum standard absolute GA spotted: adults only accepted item #1 as full; children’s
judgments, however, were less sharp. In order to assess whether children were adult-like in their judgments of *full*, we conducted a comparison of *full* to the relative GAs *big* and *long*.

If children were adult-like in their judgments, they should have demonstrated a tendency to allow the second and third items to be *big* or *long* but not allow them to be *full*. We therefore targeted the second and third item in the series for statistical analysis. Even though the children were not at floor with *full*, as adults were, for both the second and third items, the difference between the *big* and *full* judgments and between the *long* and *full* judgments is significant in one-tailed t tests (item #2, *big/full*: \( t(35) = -3.654, p < 0.0001 \); *long/full*: \( t(35) = -5.596, p = 0.000001 \); item #3, *big/full*: \( t(35) = -2.485, p = 0.009 \); *long/full*: \( t(35) = -4.448, p < 0.0001 \)). Given the S-shaped curve of the data, we also analyzed the children’s judgments of the four adjectives for each age group and for each item in the series using a best-fitting logistic regression. While the slopes for *big* and *long* judgments are not significantly different from each other (\( t(4) = 0.05, p = 0.48 \)), the slope for the *big* judgments is significantly different from those for *full* (\( t(4) = 4.59, p = 0.005 \)), as is the slope for the *long* judgments (\( t(4) = 7.28, p < 0.001 \)).

There was also an effect of condition (‘Relative-Absolute’ v. ‘Absolute-Relative’). Eleven of the fifteen children (73.3%) who judged a container other than the first one to be full were in the ‘Relative-Absolute’ condition, and therefore saw the *big* and *long* sets before the *full* and *spotted* sets. These children were scattered across age groups (three 3-year-olds, four 4-year-olds, and four 5-year-olds). Put another way, eleven of the eighteen children (or 61.1%) in the ‘Relative-Absolute’ condition judged the second container to be full, while only four of the eighteen children (or 22.2%) in the ‘Absolute-Relative’ condition did. Six children across age groups seemed to interpret *full* minimally (e.g., ‘filled to some degree’) and consequently judged the first six items in the set to be full. Excluding these children from the condition analysis, we still observe the same asymmetry between conditions: seven
children in the ‘Relative-Absolute’ condition, and only two in the ‘Absolute-Relative’ condition judged the second item to be full.

3.4 Discussion

There are three main results of the Scalar Judgment Task. First, we have evidence that both children and adults make a distinction between relative and absolute GAs, and within the latter class, between minimum and maximum standard absolute GAs. Relative GAs such as *big* and *long* have standards that evoke a partitioning of a finite set of objects ranked along an appropriate scalar continuum somewhere around the midpoint of the continuum, while absolute GAs use standards that partition the sets at the ends of the continuum. The fact that the standard for absolute GAs such as *spotted* is minimal while the standard for absolute GAs such as *full* is maximal is reflected in judgments that separate the lowest-ranked item from the rest of the series for *spotted* and the highest-ranked item from the rest of the series for *full*.

Second, judgments of relative size along a scale appear to become more categorical with development. This trend is evident in the comparison of the percentages of the third and fourth items of the relative GA sets for the two age groups. The judgments are therefore similar in form to those reported by Smith, Cooney, and McCord (1986), who found that adults had broader categories with steeper slopes than children did for objects described as *big* or *long*. Finally, most children, excluding a small subset who interpreted *full* as ‘filled to some degree’, share with adults the meaning of *full* as ‘maximally filled’. This finding was supported by a comparison of judgments for the *full* set with the two relative GA sets, targeting key items in the series and the slopes of the distributions. However, we also found that children’s interpretation of *full* can be influenced by prior context, as seen in the ordering effect between conditions.
4. Experiment 1

4.1 Experiment 1a

4.1.1 Introduction

The goal of Experiment 1 was to use definite descriptions of the sort discussed in sections 1.2 and 1.3 to probe children’s sensitivity to the kinds of context/meaning interactions discussed in the introduction to this paper. 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 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). In addition, 24 adult native speakers of English (Northwestern undergraduates fulfilling an experimental requirement for a Linguistics course) served as controls.

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 test session consisted of the same materials for each participant and included target stimuli (corresponding to target relative and absolute GAs) and control stimuli. The complete set of materials is outlined in the following three tables. Table 2 presents the four pairs used in the training session. Table 3 presents the target stimuli used in the test session, while
Table 4 presents the control stimuli used in the test session. The column on the far right in each table corresponds to an experimental feature discussed in the Procedure section.

Table 2: Training stimuli for Experiment 1

<table>
<thead>
<tr>
<th>adjective</th>
<th>stimuli</th>
<th>pragmatic status of request</th>
</tr>
</thead>
<tbody>
<tr>
<td>happy</td>
<td>pictures of a happy face and an angry face</td>
<td>felicitous</td>
</tr>
<tr>
<td>round</td>
<td>pictures of a green triangle and a blue square</td>
<td>infelicitous</td>
</tr>
<tr>
<td>red</td>
<td>pictures of a red circle and a red square</td>
<td>infelicitous</td>
</tr>
<tr>
<td>blue</td>
<td>pictures of a yellow bird and a blue bird</td>
<td>felicitous</td>
</tr>
</tbody>
</table>
Table 3: Target stimuli for Experiment 1

<table>
<thead>
<tr>
<th>adjective</th>
<th>stimuli</th>
<th>pragmatic status of request</th>
</tr>
</thead>
<tbody>
<tr>
<td>relative</td>
<td></td>
<td></td>
</tr>
<tr>
<td>big</td>
<td>two blocks, one bigger than the other</td>
<td>felicitous*</td>
</tr>
<tr>
<td></td>
<td>(blocks #1 and 3 from SJT)</td>
<td></td>
</tr>
<tr>
<td>big</td>
<td>two blocks, one bigger than the other,</td>
<td>felicitous*</td>
</tr>
<tr>
<td></td>
<td>both smaller than the first pair</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(blocks #5 and 7 from SJT)</td>
<td></td>
</tr>
<tr>
<td>long</td>
<td>two rods, one longer than the other</td>
<td>felicitous*</td>
</tr>
<tr>
<td></td>
<td>(rods #1 and 3 from SJT)</td>
<td></td>
</tr>
<tr>
<td>long</td>
<td>two rods, one longer than the other,</td>
<td>felicitous*</td>
</tr>
<tr>
<td></td>
<td>both shorter than the first pair</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(rods #5 and 7 from SJT)</td>
<td></td>
</tr>
<tr>
<td>absolute</td>
<td></td>
<td></td>
</tr>
<tr>
<td>spotted</td>
<td>two disks, one with a few spots,</td>
<td>felicitous</td>
</tr>
<tr>
<td></td>
<td>one without any</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(disks #5 and 7 from SJT)</td>
<td></td>
</tr>
<tr>
<td>spotted</td>
<td>two disks, one more spotted than the other</td>
<td>infelicitous</td>
</tr>
<tr>
<td></td>
<td>(disks #1 and 4 from SJT)</td>
<td></td>
</tr>
<tr>
<td>full</td>
<td>full container and one filled more than halfway</td>
<td>felicitous</td>
</tr>
<tr>
<td></td>
<td>(containers #1 and 3 from SJT)</td>
<td></td>
</tr>
<tr>
<td>full</td>
<td>two containers filled somewhat,</td>
<td>infelicitous</td>
</tr>
<tr>
<td></td>
<td>one more than the other, but neither full</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(containers #4 and 6 from SJT)</td>
<td></td>
</tr>
</tbody>
</table>
Table 4: Control stimuli for Experiment 1

<table>
<thead>
<tr>
<th>adjective</th>
<th>stimuli</th>
<th>pragmatic status of request</th>
</tr>
</thead>
<tbody>
<tr>
<td>color</td>
<td></td>
<td></td>
</tr>
<tr>
<td>yellow</td>
<td>pictures of a yellow bird and a black bird</td>
<td>infelicitous</td>
</tr>
<tr>
<td>green</td>
<td>purple yo-yo and yellow yo-yo</td>
<td>infelicitous</td>
</tr>
<tr>
<td>red</td>
<td>pictures of a red square and a red circle</td>
<td>infelicitous</td>
</tr>
<tr>
<td>red</td>
<td>red poker chip and a white poker chip</td>
<td>felicitous</td>
</tr>
<tr>
<td>shape</td>
<td></td>
<td></td>
</tr>
<tr>
<td>round</td>
<td>pictures of a red triangle and a red square</td>
<td>infelicitous</td>
</tr>
<tr>
<td>square</td>
<td>pictures of a blue square and a yellow circle</td>
<td>felicitous</td>
</tr>
<tr>
<td>mood</td>
<td></td>
<td></td>
</tr>
<tr>
<td>happy</td>
<td>pictures of a sad face and an angry face</td>
<td>infelicitous</td>
</tr>
<tr>
<td>happy</td>
<td>pictures of an angry face and a sad face</td>
<td>infelicitous</td>
</tr>
<tr>
<td>sad</td>
<td>pictures of a happy face and a sad face</td>
<td>felicitous</td>
</tr>
</tbody>
</table>

Procedure

Participants were invited to play a game. Children were introduced to a puppet (played by a second experimenter) and were told 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. Adult participants interacted with one 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 & Gelman, 1988; Eilers, Kimbrough Oller, & Ellington, 1974; Gelman & Ebeling, 1989; Gelman & Markman, 1985; Harris, Morris, & Terwogt 1986; Ravn & Gelman, 1984; Sena & 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 noun phrase of the form *the A one*, where *A* was an adjective (e.g., *Please give me the red one*). 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). 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)*.

Pairs of objects were designed so that they either satisfied or violated one or both of the presuppositions of the definite description. (See the far right column of the three previous tables.) 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 a red object and a white object). 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 because neither member of the pair fit the description (e.g., there was a yellow object and a blue object).

For our crucial pairs – those involving relative GAs and marked as ‘felicitous*’ in Tables 3 and 5 – whether the request was felicitous or not depended on how the adjectival modifier in the description was interpreted. If it was interpreted in the same way as in the Scalar Judgment Task (e.g., if the judgments about the size or length of items #1, 3, 5, and 7 remained constant across tasks), then either an existence or uniqueness violation would be
incurred, depending on whether the two objects were from the lower part of the continuum or the upper part, respectively. If, however, a new standard of comparison (distinct from the one used in the Scalar Judgment Task) was posited in order to ensure that the adjective would be true of just one object (the bigger or longer one; see the discussion of this point in section 1.2), the request would be felicitous. If participants accepted requests involving relative GAs, then, this would constitute evidence both that they know that such GAs are context sensitive in the way outlined in section 1.2, and (given the responses in the Scalar Judgment Task) that they are sensitive to, and willing to accommodate, the presuppositions of a definite NP.

Pairs involving absolute GAs provided crucial controls and substantiation of this conclusion. Since 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 involve a violation of the uniqueness requirement of the definite description, 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, because the request violates the existence presupposition of the definite description, as there is no full container.

If on the other hand *spotted* and *full* were context dependent in the same way as *big* and *long*, participants ought to be able to shift their meanings in a way that accommodates the presuppositions of the definite description in these examples, as with the relative GAs. Rejection of presupposition-violating requests involving absolute GA pairs, coupled with acceptance of requests involving relative GA pairs, would therefore constitute evidence for the relative/absolute distinction as laid out here, and would show that participants are, in fact,
modulating the interpretation of relative GAs in accord with the presuppositions of the definite NP.

At the same time, this pattern of results would provide important evidence that participants are not treating the adjective in the definite description as semantically equivalent to the comparative form (i.e., 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 would 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 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 & Wales, 1970; Ehri, 1976; Finch-Williams, 1981; Gathercole, 1979; Gitterman & Johnston, 1983; Graziano-King, 1999; Graziano-King & Cairns, 2005; Layton & Stick, 1978; Moore, 1999). If participants reject presupposition-violating requests involving absolute GAs, however, we can be confident that they are not reanalyzing the adjectives as comparatives.

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. (See Table 2.) Once it was evident that participants felt comfortable with the task, we proceeded with 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.

Items in the test session included target items whose salient property corresponded to one of the target GAs, and control items. 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, a point that becomes important in the interpretation of the results. Specific details about the request, stimuli, and predicted response for the target GA items are outlined in Table 5.