Journal of Semantics 27: 1–35
doi:10.1093/jos/ffp011
Advance Access publication December 3, 2009

Meaning and Context in Children's Understanding of Gradable Adjectives

KRISTEN SYRETT
Rutgers University

CHRISTOPHER KENNEDY
University of Chicago

JEFFREY LIDZ
University of Maryland

Abstract

This paper explores what children and adults know about three specific ways that meaning and context interact: the interpretation of expressions whose extensions vary in different contexts (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). The empirical focus is the use of unmodified (positive form) gradable adjectives (GAs) in definite descriptions to distinguish between two objects that differ in the degree to which they possess the property named by the adjective. We show that by 3 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 GA appearing in it. These findings suggest that children's semantic representations of the GAs 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 variability', specifically semantic context dependence v. pragmatic tolerance of imprecision.

1 INTRODUCTION

1.1 *The goals of this study*

This study investigates children's behaviour in evaluating a particular type of expression—definite descriptions based on gradable adjectives (GAs)—to investigate the nature of children's understanding of contextual aspects of meaning, and as a means of distinguishing

between different varieties of context sensitivity. As part of this investigation, we also provide experimental support for a distinction between two classes of gradable predicates that differ in whether they express relations to context-dependent (relative) or fixed (absolute) standards of comparison in the positive form. In the remainder of this section, we present the theoretical and empirical background for the study. The experiments are presented in sections 2–4. Section 5 details the theoretical implications of the experimental results.

1.2 *Relative GAs and standards of comparison*

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). 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*), extra-linguistic knowledge (e.g. that a snowman described as 'tall' was built by third graders v. fraternity brothers; see Kamp & Partee 1995), and the interests and expectations of the participants in the discourse (Graff 2000).

In what follows, we will refer to GAs that have context-dependent positive forms as *relative* GAs (to be distinguished later in section 1.4 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 & Vennemann 1972; Wheeler 1972; Cresswell 1976; Klein 1980 1991; von Stechow 1984; Ludlow 1989; Kennedy 1999 2007; *inter alia*). The standard of comparison is a degree that corresponds to something like 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 GAs 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 GA 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 & Vennemann 1972).[1]

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

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) $[\![pos\ \mathrm{tall}]\!]^c = \lambda x . \mathrm{tall}(x) \geqslant \mathrm{s}(\mathrm{tall})(c)$

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: $c(gymnasts)$, in which we are talking about female gymnasts, and $c(women)$, 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 $c(gymnasts)$ is lower than the standard in $c(women)$, 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 $c(gymnasts)$ and $c(women)$, as represented graphically in (3), then given (1), Anna is in the extension of $[\![pos\ \mathrm{tall}]\!]^{c(gymnasts)}$ but is not in the extension of $[\![pos\ \mathrm{tall}]\!]^{c(women)}$, and the truth of *Anna is tall* 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).

(3) HEIGHT: 0 - s(tall)(*c*(*gymnasts*)) - tall(anna) - s(tall)(*c*(*women*)) ->

In what follows, we will generally omit *pos* when referring to the positive form for perspicuity (i.e. we will say *tall* rather than *pos tall*), unless it is important to distinguish the compositional details of the positive form from some other form.

1.3 *Shifting standards in definite descriptions*

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: use of a singular definite determiner phrase (DP) of the form *the* ϕ introduces 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 Russell 1905; Strawson 1950; Heim 1990; Kadmon 1990; Neale 1990; Birner & Ward 1994; Abbott 1999; Roberts 2003; 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 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 existence presupposition is not met: there is no object that satisfies the property *red rod* in the context. (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 uniqueness presupposition of the definite description

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

the blue rod is not met: there are two objects in the context that satisfy the property *blue rod*. Speaker A can, however, felicitously use (4c) to request the longer of the two rods.

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

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) $\lambda x.\text{long}(x) \geqslant \text{s}(\text{long})(c)$

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—that is, independently of the prevailing standard of length (for rods). This is thus an instance of

[3] This is a bit of an oversimplification. As pointed out in Kennedy (2007: 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 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. Kamp 1980; Bosch 1983; Raffman 1994, 1996; Kamp & Partee 1995; Soames 1999; Graff 2000), but since our focus here is on relatively clear cases of context-dependent shifts in extension, we will set aside questions about vagueness.

presupposition accommodation (Lewis 1968, 1970), whereby a new standard is accommodated in order to make the discourse model consistent with the presuppositions of the definite description.[4]

1.4 *Absolute GAs*

It has often been assumed that all GAs have context-dependent positive forms, and some influential analyses of gradability and comparison have been built on this premise (Lewis 1970; Kamp 1975; Klein 1980; van Rooij forthcoming). However, 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 Kennedy & McNally (2005) and Unger (1975), we refer to this class as *absolute* GAs (see also Rusiecki 1985; Rotstein & Winter 2004; Kennedy 2007).[5] Previous researchers (see Rusiecki 1985; Cruse 1986; Rotstein & Winter 2004; Kennedy & McNally 2005; Kennedy 2007; Syrett 2007) have argued for a distinction between absolute and relative GAs on the basis of modifier selection and entailment patterns 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, *minimum standard absolute* GAs such as *spotted*, *bumpy*, *wet*, *bent* and *impure* merely require their arguments to merely possess some non-zero 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 relative GAs; on our analysis, this means that they denote measure functions. We can further account for the fixed standards of the positive form by positing denotations like (7a–b) for maximum and minimum

[4] While accommodation of a standard in definite descriptions is in general quite flexible, some adjectives do impose additional constraints. For example, evaluative adjectives such as *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 (see Bierwisch 1989; Rett 2008).

[5] Rotstein & Winter (2004) refer to the two classes of absolute GAs we discuss (maximum and minimum standard) as *total* and *partial* gradable predicates, respectively.

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 of the ranges of the measure functions denoted by the adjectives (the scales).

(7) a. $\lambda x.\text{full}(x) = \max(\text{full})$ (maximum standard GA)
b. $\lambda x.\text{spotted}(x) > \min(\text{spotted})$ (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 truth conditions equivalent to (7a–b) are actually derived, but rather on the predictions about the meaning and acceptability of 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*).[6]

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 patterns of definite descriptions discussed in the previous section, which involved presupposition accommodation via contextual shifting of a standard of comparison (cf. (4)). Suppose we discover that children correctly associate, for example, *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, that they are misanalysing the morphologically positive *long* as the implicit semantically comparative *longer* (cf. work by Nelson & Benedict 1974). If this were the case, we

[6] Note, however, that we maintain the core assumption that all GAs denote functions of type ⟨e,d⟩ (or alternatively, 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 & Winter 2004; Kennedy & McNally 2005). For example, composition of *full* with a comparative [*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.

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 analyse 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*. Children should always accept a description with a relative GA as applied to the higher ranked object, as outlined above. However, children should reject a description with a maximum standard absolute GA 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, given (7a). Similarly, children should reject a description with a minimum standard absolute GA 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 disks), as this violates the uniqueness presupposition, given (7b).[7] In contrast, if children analyse 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 can 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 such as *full*. Even though the literature cited above has provided a number of arguments in favour 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 become murkier when we consider a jar that is 7/8 or 15/16 full, or any amount that is extremely close to being full. At some point we would typically be willing to start saying that the jar is full and would consequently be willing to refer to it as *the full jar*, and this point

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

might be different in different contexts (based on our goals, the types of jars, the contents of the jars, etc.).

One interpretation of these intuitions is that the denotation in (7a) is incorrect, and instead *full* should be assigned a context-dependent meaning just like *long* or *tall* (cf. Lewis 1979). However, another possibility is that (7a) 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 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).[8]

2 EXPERIMENT 1

2.1 *Introduction*

The experiments reported in this paper use definite descriptions to probe children's sensitivity to the kinds of context–meaning interactions discussed in sections 1.2–1.4. Specifically, we were interested in determining whether children correctly shift the standard of comparison for relative GAs to accommodate the existence and uniqueness presuppositions of the definite description, and correctly *avoid* doing so for absolute GAs.[9]

2.2 *Method*

2.2.1 *Participants* Thirty children representing three age groups participated in this task: 10 three-year-olds (five boys and five girls,

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

[9] 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. Clark 1972, 1973; Klatzky *et al.* 1973; Eilers *et al.* 1974; Brewer & Stone 1975; Bartlett 1976; Townsend 1976; Marschark 1977; Carey 1978; Keil & Carroll 1980; Barner & Snedeker 2008).

range: 3;5–3;11, M: 3;8); 10 four-year-olds (four boys and six girls, range: 4;1–4;11, M: 4;5) and 10 five-year-olds (three boys and seven girls, range: 5;1–5;8, M: 5;5). Twenty-four 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.

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

2.2.3 *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 could not, 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. Eilers *et al*. 1974; Bartlett 1976; Ravn & Gelman 1984; Gelman & Markman 1985; Ebeling & Gelman 1988; Gelman & Ebeling 1989; Harris *et al*. 1986; 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 DP of the form *the A one*, where *A* was an adjective (e.g. 'Please give me the red one'). As noted in section 1.3, this type of description presupposes both existence (e.g. there must be a red one) and uniqueness (e.g. there must be exactly one red one). We deliberately varied the felicity conditions of the request, presenting participants with pairs of objects that 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). 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).

Based on the semantics of the GAs discussed in the previous sections, we predict the following. Because relative GAs such as *big* and *long* depend on the context for the standard of comparison, participants should posit a new standard of comparison each time a new pair is introduced in order to ensure that the adjective is true of just one object (i.e. the bigger or longer one). Thus, participants should always be able to accommodate the presuppositions of the definite description and accept the request as felicitous. As we mentioned in the previous section, pairs involving absolute GAs provide important controls. Because 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 when they have significantly different numbers of spots, because this would incur a uniqueness violation. Likewise, if *full* means 'is maximally full', participants should reject requests for *the full one* when confronted with two partially filled containers, even when they contain significantly different amounts of stuff, because the request incurs an existence violation.

Thus, rejection of presupposition-violating requests with the absolute GA pairs, coupled with acceptance of requests with the relative GA pairs, should constitute evidence for a semantic distinction between relative and absolute GAs. At the same time, this pattern of results should also provide evidence that participants are not just treating the positive form of the adjective in the definite description as semantically equivalent to the comparative form, that is 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 & Wales 1970;

Ehri 1976; Layton & Stick 1978; Gathercole 1979; Finch-Williams 1981; Gitterman & Johnston 1983; Graziano-King 1999; Moore 1999; Graziano-King & Cairns 2005).

Acceptance of presupposition-violating requests could be interpreted in different ways, depending on the results. If such requests are accepted across the board, we must conclude that positive form absolute adjectives—and presumably relative ones as well—can be reanalysed as comparatives, for the reasons outlined above. If, however, such requests are accepted only in certain contexts—which is in fact what we will see below—we will need to consider the possibility that absolute adjectives also have standards that can be shifted in different contexts to accommodate the presuppositions of the definite determiner.

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 to 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 Tables 1 and 2. Table 1 presents the control stimuli used in the test session, while Table 2 presents the target stimuli used in the test session. Target GA stimuli were selected from four sets of adjectives used in previous experiments (Syrett *et al.* 2006) designed to identify default 'cut-off points' for relative and absolute GAs.

The presentation of the pair members was counter-balanced 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 pseudo-randomized 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.

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

Adjective	Stimuli	Pragmatic status of request
Colour		
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 (existence)
Red	Pictures of a red square and a red circle	Infelicitous (uniqueness)
Shape		
Square	Pictures of a blue square and a yellow circle	Felicitous
Round	Pictures of a red triangle and a red square	Infelicitous (existence)
Mood		
Sad	Pictures of a sad face and a happy face	Felicitous
Happy	Pictures of a sad face and an angry face (a)	Infelicitous (existence)
Happy	Pictures of a sad face and an angry face (b)	Infelicitous (existence)

Table 1 Control stimuli for Experiment 1

Adjective	Stimuli	Pragmatic status of request
Relative		
Big	Two big blocks, one bigger than the other	Felicitous
Big	Two small blocks, one bigger than the other	Felicitous
Long	Two long rods, one longer than the other	Felicitous
Long	Two short rods, 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 (uniqueness)
Full	Two jars, one full, one about 2/3 full	Felicitous
Full	Two jars, neither full, one fuller than the other	Infelicitous (existence)

Table 2 Target stimuli for Experiment 1

software (Hollich 2003). The videos were coded frame by frame, where one frame is equal to 1/30 of a second.[10] We present an analysis of children's reaction times (RTs) for Experiments 1 and 2 in the Results section of Experiment 2.

2.3 *Results*

The distribution of responses is presented in Table 3. Recall that for each pair, we predicted one of three possible responses to the

[10] 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 five frames for any of the three measurements, a third coder was brought in as a tie-breaker for that item.

			Age group							
			3 years		4 years		5 years		Adults	
			1	0/2	1	0/2	1	0/2	1	0/2
1	Control	(felicitous)	98	2	100	0	100	0	100	0
2	Control	(infelicitous)	4	90	0	98	0	100	0	100
3	*Big*	(big)	100	0	100	0	100	0	100	0
4	*Big*	(small)	90	10	90	10	100	0	96	4
5	*Long*	(long)	80	20	90	10	100	0	100	0
6	*Long*	(short)	100	0	100	0	100	0	96	4
7	*Spotted*	(felicitous)	100	0	100	0	100	0	100	0
8	*Spotted*	(infelicitous)	20	80	30	70	10	90	4	96
9	*Full*	(felicitous)	100	0	100	0	100	0	96	4
10	*Full*	(infelicitous)	60	40	70	30	70	30	12	88

Table 3 Distribution of responses in Experiment 1

The types of control and test items, corresponding to the target adjectives, are listing in the column to the right of the row numbers. Responses for each of the item types from the four participant age groups are presented in each of the columns. Cells contain the percentage of participants responding in the manner indicated for the item in question (i.e. accepting or rejecting the request).

accompanying request, based on the above felicity patterns: participants would accept the request and give one object, give neither object, or give both objects. Both the second and third responses—either giving neither object or giving both—count as rejecting the request, since both indicate a presupposition failure for the singular definite description. For this reason, these responses were collapsed and are designated as '0/2' in Table 3. 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 cut-off. During the experimental sessions, we recorded which of the two objects the participant gave in response to the request. There were only four instances of children giving the lesser degree object across all items, and only with control items. Thus, in the following table we present only those instances where participants gave the object with the greater degree of the property, designated as '1' in Table 3.

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 7 and 9). Second, participants consistently rejected

infelicitous requests for the control items (row 2) and the infelicitous *spotted* test items (row 8). Indeed, there is no statistical 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: $P = 1.00$). In rejecting the puppet's request, children often offered responses such as 'Oh, but I have TWO red ones!'; 'What red one? He should say what shape!'; 'He thinks there must be two different colours!'; or 'They're ALL spotted!' Third, despite the size of the blocks or rods, when asked for the *big* or *long* one, participants nearly always gave the object that had the greater degree of the relevant property (cf. rows 3–6). We found no significant difference in responses to the trials containing relative adjectives and the control trials (Pearson's $\chi^2 = 0.036$, $df = 2$, $P > 0.98$). Thus, participants were willing and able to shift the standard of comparison for the two relative GAs, a pattern that reflects the context-dependent denotations participants assigned to these adjectives.

Children's responses only diverged from adults' in one instance (row 10). Instead of rejecting the puppet's request when shown the infelicitous *full* pair, as adults categorically did, children gave the puppet the fuller of the two containers when asked for *the full one* [Pearson's $\chi^2 = 13.14$ (with Yates' continuity correction), $df = 1$, $P < 0.001$]. However, we have reason to believe children do not actually differ from adults in their semantic treatment of *full*.

First, 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 experiments described in Syrett *et al*. (2006), designed to identify default standards, within approximately 3 weeks time. The fuller member of the infelicitous *full* pair from the current set of experiments appeared as the fourth item in a seven-item series of containers which children were asked to characterize as 'full' or 'not full'. In that research, not one of these 18 children judged a container beyond the third one to be full, but in the current experiment, 11 of them returned this container in response to a request from the puppet for *the full one*.

Second, this pattern cannot be because the children treated all of the adjectives as semantically comparative (understanding *the full one* as *the fuller one*), since they did not give the puppet the more spotted of the two disks when shown the infelicitous *spotted* pair. Note that rows 8 and 10 of Table 3 differ significantly (Pearson's $\chi^2 = 11.47$, $df = 1$, $P < 0.001$).

We therefore sought to determine whether there was another explanation for this behaviour. Indeed, the order to which participants was assigned appears to have had an effect. Every single one of the 15

children assigned to the order in which the infelicitous *full* pair appeared earlier in the sequence than the felicitous *full* pair accepted the puppet's request for *the full one* and gave him the fuller of the two containers. Only five of the 15 children in the condition in which this pair appeared after the felicitous *full* pair responded this way. We address the implications of this finding in the following section.

2.4 *Discussion*

Experiment 1 provides initial evidence that both adults and children distinguish between relative and absolute GAs in assigning the former context-dependent denotations and the latter context-independent denotations, where context (in)dependence is manifested in the standard of comparison: variable v. maximum/minimum. The only difference between children and adults involved the maximum standard GA *full*: a significant number of children gave the puppet the fuller container in the infelicitous condition, despite the fact that many of these same children separately judged this container not to be full. One potential explanation for this behaviour is that 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 infelicitous *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*. Experiment 2 was designed to adjudicate between these two possibilities.

3 EXPERIMENT 2

3.1 *Introduction*

The goal of Experiment 2 was to identify the source of the order effect observed in Experiment 1. 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 infelicitous *full* containers in response to his request for *the full one*.

3.2 *Method*

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

3.2.2 *Materials and Procedure* The same materials and procedure were used as in Experiment 1. The only difference was in the sequence of items. In Experiment 1, the infelicitous *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 infelicitous *full* pair so that the latter preceded the former.

3.3 *Results*

The results are presented in Table 4. Because our purpose in conducting Experiment 2 was solely to investigate the order effect described at the end of Experiment 1, here we only present responses to the two pairs whose position in the sequence was swapped (the *long* pair and infelicitous *full* pair), as well as the felicitous *full* pair for comparison. All of the other responses were a replication of Experiment 1 and will not be discussed here.

The pattern of responses is highly similar to the one observed in Experiment 1, with children across age groups—but not adults—targeting the fuller member of the infelicitous *full* pair as *the full one* in response to the speaker's request [Pearson's $\chi^2 = 5.33$ (with Yates' continuity correction), $df = 2$, $P < 0.02$].[11] (Compare the bottom row of Table 4 to

		Age group							
		3 years		4 years		5 years		Adults	
		1	0/2	1	0/2	1	0/2	1	0/2
Long	(long)	100	0	100	0	100	0	100	0
Full	(felicitous)	100	0	100	0	100	0	100	0
Full	(infelicitous)	100	0	83	17	60	40	30	70

Table 4 Distribution of key responses in Experiment 2

[11] 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.

row 10 of Table 3.) 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 infelicitous *full* pair in the three orders of Experiments 1 and 2 supports this conclusion (see Table 5).

The difference in judgments between the two orders that varied the presentation of the relative pair and the infelicitous *full* pair (a comparison between columns 1 and 2) is not significant (Fisher's exact test, two-tailed probabilities: 'column 1' v. 'column 2': $P = 0.24$); however, the difference in judgments between the order in which the infelicitous *full* pair preceded the felicitous *full* pair and the order in which the felicitous *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': $P = 0.01$). Children were apparently 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 felicitous *full* pair first (column 3 in Table 5), their responses to the infelicitous *full* pair deviated significantly from those of adults (67% v. 100%). This is surprising considering the similarity in the two age groups' performance with all of the other control and target pairs. Taking these observations as a starting point, we hypothesized that their responses to the infelicitous *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. That is, when the puppet requested *the full one* in a context in which neither of the two objects uniquely satisfied the description, the children were willing to respond by handing over the object that came closest to doing so. We further hypothesized that if this were correct, it would imply some additional

	1	2	3
	Experiment 1: *long* < *#full* < *full*	Experiment 2: *#full* < *long* < *full*	Experiment 1: *full* < *long* < *#full*
Children	0	18	67
Adults	75	70	100

Table 5 Percent of time participants *rejected* the request when shown the infelicitous *full* pair
The ordering of the target pairs is indicated by '<', where '*full*' is the felicitous *full* pair, '*#full*' is the infelicitous *full* pair and *long* is the *long* pair.

reasoning about the utterance beyond the computation of its semantic content: minimally, a recognition that the description used is false of the two objects, a determination of which object comes closest to making it true and a calculation of whether to behave as though it were true.[12]

To explore this possibility, we examined the videotapes of the sessions in Experiments 1 and 2 to examine children's RTs. 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* towards this object and his/her ultimate *touch* of this object.[13] We excluded from analysis the following items: those where the children's eye movements could not be coded (e.g. if the eyes were occluded), those where the child was already looking at or touching the stimuli before the request was uttered and those where any other experimental artefact 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 analysed varies from analysis to analysis; this number is always provided in a footnote.

We then analysed two sets of RT measurements. We first looked at the two *full* pairs, asking whether children took longer to respond to the puppet's request when it involved the infelicitous *full* pair than when it involved the felicitous *full* pair. These results are presented in Figure 1.[14] 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' $t(15) = 3.47$, $P < 0.002$). The number of asterisks above the bars in this and all other figures indicates significance (*P* value) at the .01, .001 and .0001 level, respectively.

Second, we looked across GA subclasses and compared the difference between the look and the touch of the object for the

[12] We know from the order of presentation effect illustrated in Table 5, and from the systematic rejection of the infelicitous *spotted* pair, that children did not indiscriminately behave as though the description was uniquely true of one member of the pair. This implies that when they did behave in such a way, it involved some sort of additional reasoning on their part. This reasoning may have involved a sort of cost–benefit analysis: the cost of allowing something to be *full* in that context which they would not ordinarily judge as such was minimal, and the benefit was that the speaker's request could be honored.

[13] 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.

[14] We analysed the RTs for 16 children who targeted the fuller container for each pair.

infelicitous *full* pair to those for three other key pairs—the two *big* pairs and the felicitous *full* pair—in order to determine if the increase in RTs was unique to the infelicitous *full* pair. These results are presented in Figure 2.[15]

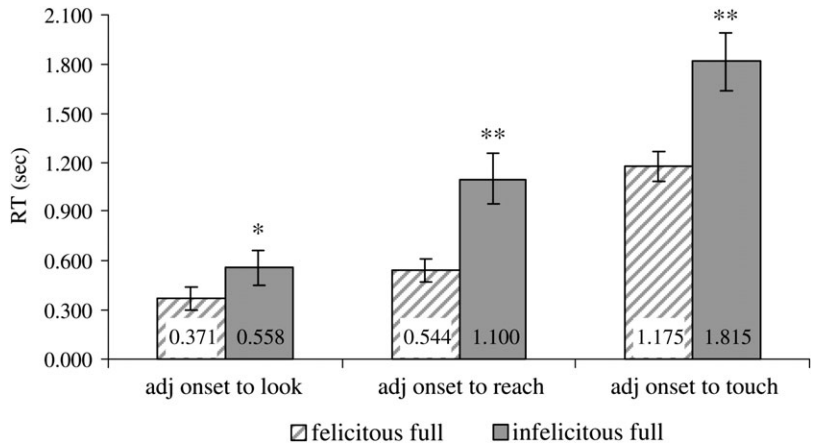


Figure 1 Children's RTs for two *full* pairs in Experiments 1 and 2.

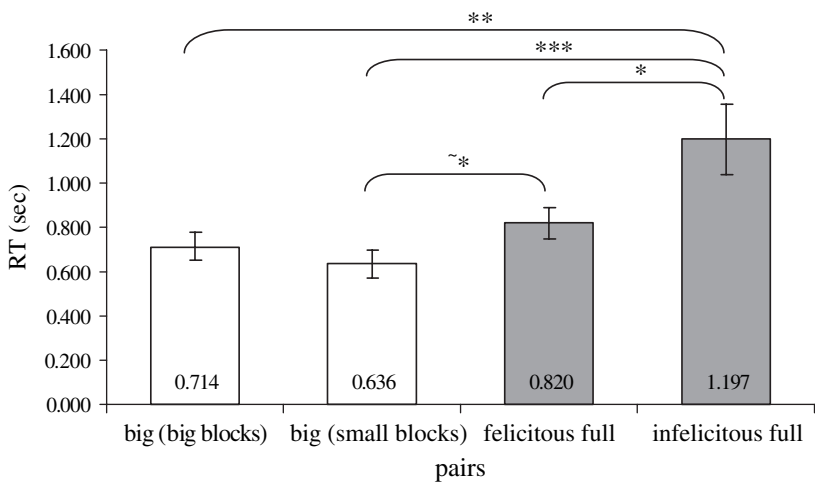


Figure 2 Children's RTs for four key pairs in Experiments 1 and 2.

[15] We analysed responses to these pairs across all children. The number of children varied for each pair: 26 for the big blocks, 32 for the small blocks, 29 for the felicitous *full* pair and 21 for the infelicitous *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.

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

These RTs demonstrate 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 render *big* false of one of two blocks both previously judged to be big, and a shift down to render *big* true of one of two blocks previously judged not to be 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*. We assess the theoretical significance of this pattern in section 5.

3.4 *Discussion*

The results of Experiment 2 elaborate upon those of Experiment 1 by demonstrating that children's decision to give the puppet the fuller of the infelicitous *full* containers in response to his request for *the full one* was not driven by the preceding *long* relative GA pair. Regardless of the order of these two target pairs, children were likely to accept the puppet's request and give him the fuller container if they had not already seen the felicitous *full* pair, which exemplifies the maximum standard. However, even with the assistance of the felicitous *full* pair, some children were still inclined to return the fuller of the infelicitous *full* pair when asked for *the full one*.

Our RT analysis shed light on this pattern by demonstrating that children took significantly longer to give the puppet the fuller of the two containers that are not full than they did 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 behaviour on the crucial items. We provide a more detailed analysis of this phenomenon in section 5. Before proceeding with that discussion, however, we first seek to determine whether this pattern of results can be extended beyond the target

absolute GAs in Experiments 1 and 2. This was the purpose of Experiment 3.

4 EXPERIMENT 3

4.1 *Introduction*

The goal of Experiment 3 was to make sure that the results in Experiments 1 and 2 are not simply facts about *full* and *spotted*.

4.2 *Method*

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

4.2.2 *Materials and Procedure* The materials and procedure 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*; in place of the two *spotted* pairs, there were two pairs corresponding to the minimum standard absolute GA *bumpy*. (See Rotstein & Winter 2004; Kennedy & McNally 2005; Kennedy 2007; and Syrett 2007 for discussion of diagnostics of absolute maximum v. minimum GAs.) These pairs were designed similarly to those in Experiments 1 and 2, so that only one of the two pairs for each adjective satisfied the presuppositions of the definite description (see Table 6).

4.3 *Results*

The results for Experiment 3 are presented in Table 7. These results can be compared with those from Experiments 1 and 2, 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 the new absolute GA pairs. All other responses were replications of Experiment 1.

Children's responses are similar to those seen in Experiment 1 with the exception of behaviour of the 4- and 5-year-olds in response to the infelicitous *straight* pair; these children were more likely to accept such requests with the infelicitous *full* pair and reject them with the infelicitous *straight* pair (Pearson's $\chi^2 = 4.9$ with Yates' continuity

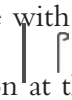
Adjective	Stimuli	Pragmatic status of request
Bumpy	Two orange wooden, 2″ × 5″ boards; One with some bumps and one with none	Felicitous
	Two orange wooden, 2″ × 5″ boards; One with some bumps and one with more	Infelicitous (uniqueness)
Straight	A straight wire and a straight wire with a curly section at the top 	Felicitous
	A straight wire with a curly section at the top and a fully curled wire 	Infelicitous (existence)

Table 6 Absolute GA stimuli used in Experiment 2

		Age group							
		3 years		4 years		5 years		Adults	
		1	0/2	1	0/2	1	0/2	1	0/2
Bumpy	(felicitous)	70	0	90	0	100	0	96	4
Bumpy	(infelicitous)	20	70	20	70	10	80	0	100
Straight	(felicitous)	90	10	90	0	90	10	100	0
Straight	(infelicitous)	70	30	40	60	20	70	12	88

Table 7 Distribution of responses in Experiment 2

	Experiment 1	Experiment 2	Experiment 3
Before felicitous maximal pair	0	18	38
After felicitous maximal pair	67	n/a	71

Table 8 Percent of time children *rejected* the puppet's request for *the full/straight one* when shown an infelicitous *full/straight* pair

correction, df = 1, $P < 0.03$). (Compare row 10 of Table 3 to the bottom row of Table 7.) Once again, however, we observed an order of presentation 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 infelicitous *straight* pair earlier in the sequence of items than the felicitous *straight* pair. A comparison of the two orders of presentation highlights the consistency of this effect across the three experiments (see Table 8). Children were more likely to reject the puppet's request when the infelicitous pair followed the felicitous pair in the sequence of items.

Turning now to RTs, although there were only five children for which the within-experiment comparison of the felicitous *straight* and infelicitous *straight* pairs could be made, the same pattern emerges (see Figure 3).[16] As with the *full* data, the differences between the adjective onset to the look, reach and touch between the two *straight* 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' $t(4) = -3.32$, $P = 0.015$).

As in Experiments 1 and 2, we also compared RTs across children, examining the difference between the look and the touch for the felicitous *straight* pair, the infelicitous *straight* pair and the two *big* pairs.

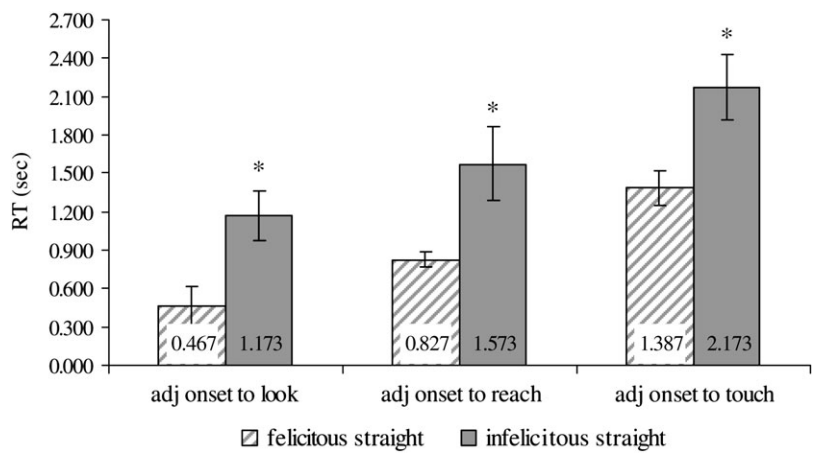


Figure 3 Children's RTs for two *straight* pairs in Experiment 2b.

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

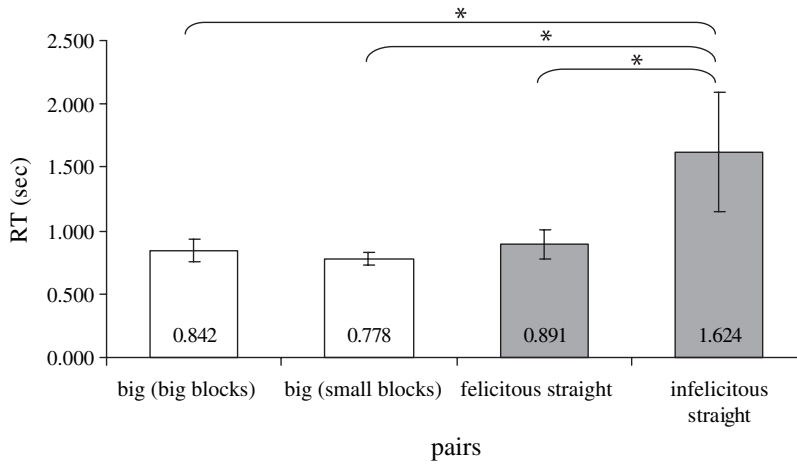


Figure 4 Children's RTs for four key pairs in Experiment 2.

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: 'infelicitous *straight*' v. 'felicitous *straight*': $t(23) = 2.15$, $P = 0.04$; v. 'big' (big blocks): $t(20) = 2.31$, $P = 0.03$; v. 'big (small blocks): $t(22) = 2.75$, $P = 0.01$). No other significant differences were found (two-tailed *t*-tests: 'big (big blocks)' v. 'big (small blocks)': $t(30) = 0.64$, $P = 0.53$; 'felicitous *straight*' v. 'big (big blocks)': $t(31) = -0.33$, $P = 0.74$; 'felicitous *straight*' v. 'big (small blocks)': $t(33) = 0.86$, $P = 0.40$).[17]

4.4 *Discussion*

The results of Experiment 3, which replicated those of Experiments 1 and 2, provide evidence that the pattern of behaviour with absolute GAs is not due to lexical idiosyncrasy, but rather reflects aspects of the meanings of these terms. The rigidity of the minimum standard of *spotted*, which results in a violation of the definite description's uniqueness presupposition in the infelicitous stimulus pair, was also observed to be a property of the minimum standard GA *bumpy*.[18] At

[17] The data from the following number of children were analysed for each pair: 15 for the big blocks, 17 for the small blocks, 18 for the felicitous *straight* pair and 7 for the infelicitous *straight* pair.

[18] This result also provides further evidence that children are not reanalysing the adjectives in the crucial stimuli as comparatives, and arguably stronger evidence than the *spotted* stimuli in Experiments 1 and 2: while *spotted*~*more spotted* is arguably not a fully colloquial alternation (we tend to say *x is has more spots than y* rather than *x is more spotted than y*), *bumpy*~*bumpier* is one.

the same time, the effect of order of presentation on participants' willingness to accept the infelicitous absolute maximum standard pair that we observed with *full*, and corresponding increase in RTs, was also seen with *straight*. This indicates that the interpretive/evaluative processes involved in choosing an object that comes closest to satisfying a description based on an absolute maximum standard GA are different from those involved in choosing an object that satisfies a context-dependent description based on a relative GA. In the next section, we propose an explanation for these differences.

5 GENERAL DISCUSSION

We began this paper by outlining three main goals: to investigate young children's understanding of the contextual aspects of meaning (as illustrated in definite descriptions based on GAs), to explore distinctions in the nature of contextual sensitivity (whether semantic or pragmatic) and to provide experimental support for a distinction between two classes of GAs based on the role of context. In this section, we summarize how our experimental results address each of these goals and present an account of the differences in RT for the absolute maximum standard pairs.

First, we have shown that by 3 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 routinely 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 encoding a mapping from objects to scalar representations, but they differ in whether the positive form denotes a relation to a context-sensitive standard of comparison or to a fixed one.[19]

[19] This difference arguably 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 Rotstein & Winter 2004; Kennedy & McNally 2005; and Kennedy 2007). If this is correct, then a more accurate way to state the generalization is that by 3 years of age, children already know the mapping between scale structure and the denotation of the positive form (Syrett 2007).

Second, children's responses in the PAT show that by 3 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. At the same time, their responses to the stimuli involving relative GAs show that when presupposition accommodation is an option—achieved 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, 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.

Finally, our results provide new data on the nature and variety of context sensitivity in meaning and use. The core behavioural responses in the PAT (whether the participants accepted the request and returned the object with the greater degree of the relevant property) showed that children treated both descriptions based on relative GAs and descriptions based on maximum standard absolute GAs as (potentially) variable: objects that otherwise would not satisfy a particular stimulus adjective *A* were nevertheless judged to be acceptable referents of a definite description of the form *the A one* when they were paired with a second object that manifested the scalar property encoded by *A* to a lesser degree. As we have said, this result cannot be attributed to reanalysis of *A* as a comparative, since parallel behaviour was not observed with the minimum standard absolute GAs. Instead, this result shows that in the context of the task, subjects were behaving as though the description was true of one object and false of the other. In the case of objects that failed to manifest the maximum standard of an absolute GAs, this result occurred less frequently than it did with relative GAs, and was influenced by the order of presentation, but it occurred often enough to indicate the influence of some sort of context sensitivity. The question is whether the same kind of context sensitivity is active in both cases.

Note that this question applies outside of the domain of this task and bears on both child and adult language. Although we did not observe the same variability in descriptions based on maximum standard GAs 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 infelicitous *full* containers had been closer to full than it actually was (though still noticeably not full), and the straighter of the two wires

closer to straight, adults would also have been inclined to provide these objects in response to the experimenter's request for *the full/straight one*. Such results would replicate the intuitions about maximum standard adjectives that we discussed at the end of section 1.4.

These intuitions are based on a simple observation about everyday linguistic behaviour: we regularly speak imprecisely. In addition to using absolute GAs such as *full* and *straight* to describe objects that are not (strictly speaking) maximally full or straight, we also say 'We arrived at 3 p.m.' to describe an arrival somewhere at 3:03 p.m., and we say 'The child ate all her dinner' when she consumed of all but a few bites of the food on her plate (Sperber & Wilson 1986; Lasersohn 1999). 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 it would be inappropriate to use 'We arrived at 3 p.m.' to describe an arrival at 3:03), one interpretation of these facts is that these expressions also have context-dependent denotations, like relative GAs. On this view (advocated by, e.g. Lewis 1970, 1979; Kamp 1975; and Pinkal 1995), interpretive variability is always fundamentally semantic, and expressions like *full* and *straight* (and perhaps *arrived at 3 p.m.*, *ate her dinner*, etc., as well) have meanings that, like *big* and *long*, require fixing the value of some contextual parameter as part of determining their extensions. Another interpretation is that all of these expressions, including relative GAs, have fixed denotations, and the observed variability is a purely pragmatic phenomenon (Austin 1979; Travis 1994, 1996).

Either of these 'uniform' accounts of contextual variability in scalar predicates would be consistent with children's pattern of object selection in the PAT. However, neither view provides a satisfactory explanation of the RT data that we collected, which resulted in two findings. First, the time between hearing a definite description containing a relative GA and making a decision about which of two objects to select based on the description was the same regardless of whether the objects were separately judged to satisfy the adjective or not (e.g. whether the blocks were judged to be big or not). Second, the time between hearing a definite description containing a maximum standard absolute GA and making a decision about whether to select the object that came closest to satisfying the adjective in the infelicitous case was significantly longer than the time it took to select the object that manifested the maximum standard in the felicitous case, and also significantly longer than the relative GA RTs.

The first result is expected. If children know the meaning of, for example (positive form) *big*, then they know that it denotes the

property of having a size that exceeds a variable, contextually determined standard of comparison. In a context in which (i) there are only two objects under consideration, (ii) there is a presupposition that only one of them has the property denoted by *big* (thanks to the use of *the*) and (iii) they manifest this property to different degrees, then selection of the object that uniquely satisfies *big* is straightforward: it is the bigger of the two objects. If *full* had a similar sort of denotation—if it meant 'fuller than a contextual standard' or 'at least as close to maximally full as a contextual standard'—then selection of the fuller of two objects in the infelicitous pair ought to follow a similar pattern. The second result (the longer RTs) shows that it does not, even though the end result (selection of the object with the greater degree of fullness) is the same. We conclude from this result that *full* does not have the same kind of denotation as *big* and that the reasoning involved in accepting 'variable' interpretations of maximum standard absolute GAs, while sensitive to contextual factors (such as whether an object manifesting the maximal standard has already been encountered, as seen in the order of presentation effects), is not the same as the reasoning involved in evaluating a relative GA in different contexts.

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 assessment of whether one of the objects was close enough to having the property denoted by the maximum standard GA 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 kind of reasoning; for concreteness, we will assume with Lasersohn (1999) that it involves (i) computation of a set of alternative denotations of an expression, ordered relative to their proximity to its actual denotation (what Lasersohn calls a *pragmatic halo*), and (ii) a decision about how much deviation from the actual denotation will be tolerated, based on features of the discourse context. In a context in which the actual denotation of an expression is strictly speaking false, an utterance of that expression counts as 'close enough to true' and acceptable for the purposes of the conversational exchange just in case there is an element among its set of tolerable alternatives that actually would have been true if it had been substituted for the denotation of the original expression.

In the examples we are concerned with, children accept, for example *the full one* as a description of the fuller of the two containers in the infelicitous pair if they decide that there is a property among the

tolerable alternatives to the actual denotation of *full* (which requires maximal fullness) that is true of that container and false of the other one (e.g. something equivalent to *5/6ths full* or *almost full* or *nearly full*). Whether their decision is a positive one depends on the context: the order of presentation effect shows that if they have already seen an object manifesting maximal fullness, the set of tolerable alternatives shrinks, leading them to reject the request. Crucially, evaluating the request involves layers of reasoning that go beyond the computation of semantic content, even semantic content that requires fixing contextual parameters, and takes into account alternative denotations and judgments of communicative intent. This extra reasoning imposes a processing load that, we claim, results in the longer RTs we observed.

If our interpretation of the facts is correct, then we have experimental evidence for a distinction between two types of interpretive variability. One type, 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, exhibited by imprecise uses of maximum standard absolute GAs, is fundamentally *pragmatic* and involves computation of a set of alternative denotations and a judgment about which 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 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 a maximum standard should result in adults behaving like children: they should accept false descriptions based on absolute maximum standard GAs, but should take longer to do so relative to decisions based on true descriptions or on descriptions with relative GAs. More generally, if the same kind of reasoning is involved in other cases of 'loose talk', we expect to find similar delays in interpretation or evaluation, as compared to

[20] One potential objection to this interpretation of the results is that pre-schoolers are generally not imprecise, and can in fact be quite literal at times. (We thank A. Giannakidou, personal communication, 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.

(appropriately parallel) constructions in which variability is based on semantic content alone.

Acknowledgements

This work was made possible by funding from National Science Foundation (NSF) grant BCS-0094263 to Christopher Kennedy, NSF grant BCS-0418309 and National Institutes of Health (NIH) grant DC006829 to Jeffrey Lidz, NIH grant HD30410 for the Project on Child Development to Sandra Waxman and a Northwestern University Presidential Fellowship to Kristen Syrett. We are grateful to the children and staff at Warren W. Cherry Preschool, Chiaravalle Montessori, Northbrook Community Nursery School, and the School for Little Children, and the parents and children who came to our laboratory to participate. Indispensable help running subjects and/or coding data was provided by Evan Bradley, Stefanie Brody, Jessica Mershon, Carol Sweeney, Jessica Bobula, Anne Gooch, Shyaam Ramkumar, Nadia Pierrehumbert, Erin Leddon, Josh Viau and Elisa Sneed-German. We benefited from discussions with Sandra Waxman, Louise McNally, Anastasia Giannakidou and members of the Northwestern Linguistics-Psychology Acquisition Lab, University of Chicago Semantics Lab, and feedback from the audiences at GALANA 2004, the 2005 Linguistic Society of America Annual Meeting, the 2006 CUNY Human Sentence Processing Conference, the 2006 Regional Meeting of the Chicago Linguistic Society and the 2006 Chicago Scalar Meaning Workshop. Finally, we would like to thank two anonymous reviewers and Ira Noveck for detailed and very helpful comments on earlier drafts of the paper.

KRISTEN SYRETT
Rutgers Center for Cognitive Science
Psychology Building Addition, Busch Campus
Rutgers University, New Brunswick
152 Frelinghuysen Road
Piscataway, NJ 08854-8020
USA
e-mail: k-syrett@ruccs.rutgers.edu

CHRISTOPHER KENNEDY
Department of Linguistics
University of Chicago
1010 East 59th Street
Chicago, IL 60637
USA
e-mail: ck@uchicago.edu

JEFFREY LIDZ
Department of Linguistics
University of Maryland
1401 Marie Mount Hall
College Park, MD 20742-7505
USA
e-mail: jlidz@umd.edu

REFERENCES

Abbott, B. (1999). 'Support for a unique theory of definite descriptions'. In T. Matthews & D. Strolovitch (eds.), *Proceedings from Semantics and*

Linguistic Theory IX. CLC Publications. Ithaca, NY. 1–15.

Austin, J. L. (1979). *Philosophical Papers*. Oxford University Press. Oxford.

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:**205–19.

Bartsch, R & T. Vennemann. (1972). 'The grammar of relative adjectives and comparison'. *Linguistische Berichte* **20:**19–32.

Bierwisch, M. (1989). 'The semantics of gradation'. In M. Bierwisch & E. Lang (eds.), *Dimensional Adjectives*. Springer. Berlin. 71–261.

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

Bosch, P. (1983). '"Vagueness" is context-dependence. A solution to the Sorites Paradox'. In T. Ballmer & M. Pinkal (eds.), *Approaching Vagueness*. North Holland. Amsterdam. 189–210.

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, & G. Miller (eds.), *Linguistic Theory and Psychological Reality*. MIT Press. Cambridge, MA. 264–93.

Clark, E. V. (1972). 'On the child's acquisition of antonyms in two semantic fields'. *Journal of Verbal Learning and Verbal Behavior* **11:**750–8.

Clark, H. (1973). 'Space, time, semantics, and the child'. In T. Moore (ed.), *Cognitive Development and the Acquisition of Language*. Academic Press. New York. 27–63.

Cresswell, M. J. (1976). 'The semantics of degree'. In B. Partee (ed.), *Montague Grammar*. Academic Press. New York. 261–92.

Cruse, D. (1986). *Lexical Semantics*. Cambridge University Press. Cambridge.

Donaldson, M & R. J. Wales. (1970). 'On the acquisition of some relational terms'. In J. R. Hayes & R. Brown (eds.), *Cognition and the Development of Language*. John Wiley. New York. 235–68.

Ebeling, K. S. & S. A. Gelman. (1988). 'Coordination of size standards by young children'. *Child Development* **59:**888–96.

Ehri, L. (1976). 'Comprehension and production of adjectives and seriation'. *Journal of Child Language* **3:**369–84.

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.

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, Department 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, Department 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:**920–32.

Gelman, S & E. M. Markman. (1985). 'Implicit contrast in adjectives vs. nouns: implications for word-learning in preschoolers'. *Journal of Child Language* **6:**125–43.

Gitterman, D & J. R. Johnston. (1983). 'Talking about comparisons: a study of young children's comparative adjective usage'. *Journal of Child Language* **10:**605–21.

Graff, D. (2000). 'Shifting sands: an interest-relative theory of vagueness'. *Philosophical Topics* **28:**45–81.

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.

Graziano-King, J & H. S. Cairns. (2005). 'Acquisition of English comparative adjectives'. *Journal of Child Language* **2:**345–73.

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:**335–52.

Heim, I. (1990). 'E-type pronouns and donkey anaphora'. *Linguistics and Philosophy* **13:**137–77.

Hollich, G. (2003). 'SuperCoder [Computerprogram]'. (http://hincapie.psych.purdue.edu/Splitscreen/SuperCoder.dmg).

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

Kamp, H. (1975). 'Two theories of adjectives'. In E. Keenan (ed.), *Formal Semantics of Natural Language*. Cambridge University Press. Cambridge. 123–55.

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:**129–91.

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:**21–8.

Kennedy, C. (1999). *Projecting the Adjective: The Syntax and Semantics of Gradability and Comparison*. Garland. New York. (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:**345–81.

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 & D. Wunderlich (eds.), *Semantik: Ein Internationales Handbuch der Zeitgen Ösischen Forschung*. Walter de Gruyter. Berlin. 673–91.

Kyburg, A & M. Morreau. (2000). 'Fitting words: vague language in context'. *Linguistics and Philosophy* **23:**577–97.

Lasersohn, P. (1999). 'Pragmatics halos'. *Language* **75:**522–51.

Layton, T. L. & S. L. Stick. (1978). 'Comprehension and production of comparatives and superlatives'. *Journal of Child Language* **6:**511–27.

Lewis, D. (1968). 'Counterpart theory and quantified modal logic'. *Journal of Philosophy* **65:**113–26.

Lewis, D. (1970). 'General semantics'. *Synthese* **22:**18–67.

Lewis, D. (1979). 'Scorekeeping in a language game'. *Journal of Philosophical Logic* **8:**339–59.

Ludlow, P. (1989). 'Implicit comparison classes'. *Linguistics and Philosophy* **12:**519–33.

Marschark, M. (1977). 'Lexical marking and the acquisition of relational size concepts'. *Child Development* **48:** 1049–51.

Moore, D. (1999). 'Comparatives and superlatives: lexical before functional'. In A. Greenhill, H. Littlefield, & C. Tano (eds.), *Proceedings of the 23rd Annual Boston University Conference on Language Development*. Cascadilla Press. Somerville, MA. 474–81.

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

Nelson, K & H. Benedict. (1974). 'The comprehension of relative, absolute, and contrastive adjectives by young children'. *Journal of Psycholinguistic Research* **3:**333–42.

Pinkal, M. (1995). *Logic and Lexicon*. Kluwer. Dordrecht.

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

Raffman, D. (1996). 'Vagueness and context relativity'. *Philosophical Studies* **81:**175–92.

Ravn, K. E. & S. Gelman. (1984). 'Rule usage in children's understanding of "big" and "little"'. *Child Development* **55:**2141–50.

Rett, J. (2008). *Degree Modification in Natural Language*. Doctoral dissertation. Rutgers University. New Brunswick, NJ.

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:**259–88.

Rusiecki, J. (1985). *Adjectives and Comparison in English. Longman Linguistics Library*. vol. 31. Longman Group Limited London.

Russell, B. (1905). 'On denoting'. *Mind* **14:**479–93.

Sena, R & L. B. Smith. (1990). 'New evidence on the development of the word "big"'. *Child Development* **61:**1034–52.

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

Sperber, D & D. Wilson. (1986). 'Loose talk'. *Proceedings of the Aristotelian Society New Series* **86:**153–71.

Strawson, P. F. (1950). 'On referring'. *Mind* **59:**320–44.

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

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, & B.D. Schwartz (eds.), *Proceedings of the Inaugural Conference on Generative Approaches to Language Acquisition—North America. UConn Occasional Papers in Linguistics*. vol. 4. Cambridge, MA. 353–64.

Travis, C. (1994). 'On constraints of generality'. *Proceedings of the Aristotelian Society* **94:**165–88.

Travis, C. (1996). 'Meaning's role in truth'. *Mind* **105:**451–66.

Townsend, D. (1976). 'Do children interpret "marked" comparative adjectives as their opposites?'. *Journal of Child Language* **3:**385–96.

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

van Rooij, R. (2008). 'Semi-orders and vagueness'. In P. Egré & N. Klinedinst (eds.), *Vagueness and Language Use*. Palgrave.

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

Wheeler, S. (1972). 'Attributives and their modifiers'. *Noûs* **6:**310–34.

First version received: 12.10.2007
Second version received: 20.01.2009
Accepted: 03.06.2009