Child Development, May/June 2008, Volume 79, Number 3, Pages 594 – 608

Compositionality and Statistics in Adjective Acquisition: 4-Year-Olds Interpret *Tall* and *Short* Based on the Size Distributions of Novel Noun Referents

David Barner
University of Toronto

Jesse Snedeker
Harvard University

Four experiments investigated 4-year-olds' understanding of adjective – noun compositionality and their sensitivity to statistics when interpreting scalar adjectives. In Experiments 1 and 2, children selected *tall* and *short* items from 9 novel objects called *pimwits* (1 – 9 in. in height) or from this array plus 4 taller or shorter distractor objects of the same kind. Changing the height distributions of the sets shifted children's *tall* and *short* judgments. However, when distractors differed in name and surface features from targets, in Experiment 3, judgments did not shift. In Experiment 4, dissimilar distractors did affect judgments when they received the same name as targets. It is concluded that 4-year-olds deploy a compositional semantics that is sensitive to statistics and mediated by linguistic labels.

The capacity to comprehend and generate novel utterances—and thus to use language creatively—depends on knowing that the meanings of complex expressions are a function of their syntax and the meanings of their constituent parts. Natural language is, at its core, compositional (Frege, 1892). The compositionality of language places important constraints not only on theories of concepts and their relation to linguistic structure (Fodor & Lepore, 2002) but also on theories of how children acquire language. However, although syntactic development has long been a central topic in language acquisition, few studies have directly explored the emergence of compositional semantics, to determine whether children use the same principles as adults to guide the interpretation of new word combinations. This reflects both the empirical difficulties of studying meaning (as opposed to form) and the lack of a clearly articulated alternate hypothesis. Children produce and comprehend multiword utterances from a young age. How could they do this if they were not guided by principles of semantic composition?

Recent proposals of syntactic development highlight the limitations of this logic. Tomasello and colleagues argue that although young children's early utterances appear to respect the syntax of the adult language, they are actually generated by a grammar that relies on simpler, item-based, generalizations (Lieven, Behrens, Speares, & Tomasello, 2003; Olguin & Tomasello, 1993; Tomasello, 1992, 2000). Adult-like production (or comprehension) of utterances may not necessarily implicate an adult-like representational system. Similarly, the apparent compositional nature of young children's grammar could reflect alternate strategies that result in interpretations similar enough to the target grammar to evade detection. For example, children might initially learn the meaning of complex expressions directly via ostension (e.g., "tall tree" learned in the context of tall trees) and store them whole. Such children might easily point out tall trees when asked. However, if their interpretation of *tall tree* is based on an association between that particular noun phrase and trees of a certain size, they might be incapable of interpreting *tall* in the context of other nouns, not previously paired with *tall*.

Adjectives can restrict denotation in a number of ways. Intersective adjectives (e.g., *bumpy*, *green*, *Californian*) combine with nouns to identify the

Thank you to Kristen Syrett, Katherine Chow, and an anonymous reviewer for her helpful comments on this manuscript. Thanks also to Alan Bale for his stimulating discussions, to Lorraine Barner and Jeremy Hartman for help with recruitment and data collection, to participants and caregivers from the Comox Valley, the greater Boston area, and the greater Toronto area, and to members of the Laboratory for Developmental Studies at Harvard for their feedback early in the project. Thank you also to Phil Desenne (for stimulus creation and "Pimwit ale") and Julie Burelle (for coining the term *pimwit*). The research was supported by a Harvard University Graduate Society Dissertation Completion Fellowship to David Barner and a Harvard University Mind, Brain, and Behavior Graduate Award to David Barner.

Correspondence concerning this article should be addressed to David Barner, Department of Psychology, University of Toronto, 4020 Sidney Smith Hall, 100 St. George Street, Toronto, ON M5S 3G3, Canada. Electronic mail may be sent to david.barner@utoronto.ca.

 0009-3920/2008/7903-0007

intersection of two independently identifiable sets of things. For example, *pet fish* picks out the intersection of pets and fish (‖pet‖ ∩ ‖fish‖; see Kamp & Partee, 1995), as depicted in Figure 1.

Numerous studies have investigated children's understanding of intersective adjectives to determine how they use syntactic cues to distinguish adjectives and nouns, when they use language to distinguish properties from kinds, and how kind information restricts the possible meanings of adjectives (Gelman & Markman, 1985; Hall, Waxman, & Hurwitz, 1993; Klibanoff & Waxman, 2000; Mintz, 2005; Mintz & Gleitman, 2002; Waxman & Booth, 2001). However, the relevance of these studies to evaluating children's understanding of compositionality is potentially limited. Studies demonstrating effects of syntax on novel word interpretation illustrate that children have made mappings between syntax and semantics, but they do not necessarily tell us how children combine the meanings of words to derive the meanings of phrases. The most relevant studies ask children to interpret novel adjectives in unattested adjective–noun combinations. However, even these studies are potentially ambiguous. Although children might use compositional rules to interpret novel combinations of intersective adjectives and nouns, a noncompositional strategy would also suffice.

For example, previous studies report that children 5 years of age and older interpret adjective–noun combinations incrementally, restricting the denotation of the expression as each word is heard rather than combining the words to derive a composed meaning. Matthei (1982) asked 4- and 5-year-old children to find the "second green ball" in an array in which the second ball was green but was not the second green ball. Surprisingly, children failed to pick the correct object and instead picked the second thing, which was a ball and which was green. As Hamburger and Crain (1984) later demonstrated, this observation alone does not show that children lacked access to a compositional semantics (because 5- and 6-year-olds succeed when their eyes are covered, preventing them from applying each word to the array as it is heard). However, Matthei's study does show that behaviors that are consistent with a compositional semantics may in fact reflect a simpler strategy of restricting the denotation of an expression as each individual word is heard: In contexts in which the second ball is the second green ball, children's incremental strategy would be indistinguishable from one that draws on compositional representations. This is generally true of intersective adjectives (e.g., *green*, *bumpy*, *pet*) and thus these adjectives do not allow us to easily diagnose compositionality in early child language.

A classic case for exploring compositionality in natural language is that of subsective adjectives (Partee, 1995). Subsective adjectives provide a better test of compositional knowledge because they are always interpreted relative to the noun with which they occur. These include adjectives like *tall*, *short*, *big*, *small*, *quiet*, and *loud*. In all cases, the interpretation of the adjective clearly depends on the noun. A tall boy is tall for a boy (but not for a tree). A loud frog is loud for a frog (but not for a jet). Thus, to interpret these expressions, it is not sufficient to find the intersection of the adjective and the noun or to restrict their denotation as each word is encountered in sequence; only the compositional relation between the words can define their interpretation. As evidence of this, inferences like that in (1) are not valid, whereas equivalent inferences that feature intersective adjectives are valid, as in (2; see Kamp & Partee, 1995; Partee, 1995):

1. Goldie is a big fish; Goldie is a pet; therefore, Goldie is a big pet.
2. Goldie is a black fish; Goldie is a pet; therefore, Goldie is a black pet.

Instead, subsective adjectives identify subsets of noun denotations such that *big fish* denotes the subset of fish that are big (‖big fish‖ ⊆ ‖fish‖). This is depicted in Figure 2.

For each complex concept in which they participate, subsective adjectives establish a unique criterion for identifying exemplars, typically called a "standard of comparison." For example, to determine whether

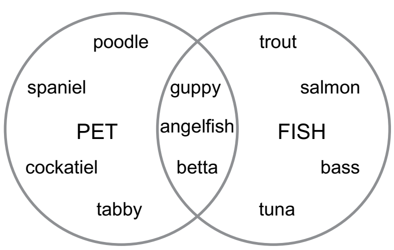

Figure 1. At the intersection of pets and fish are pet fish.

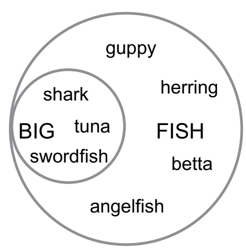

Figure 2. "Big fish" are a subset of "fish."

Goldie is a big fish, we need to know the size of Goldie and the sizes of other fish. However, to know whether Goldie is a big pet, knowing the sizes of fish is not enough. Instead, we need to know something about the sizes of pets. In each case, the semantics of the separate simple concepts and their respective identification criteria are insufficient for identifying referents of the complex concept formed by the subsective adjective–noun combination. Instead, the simple concepts must be combined to generate a composed meaning. Consequently, subsective adjectives offer an ideal test case for evaluating children's understanding of compositionality.

Gradable Adjectives and World Knowledge

What do children need to know to deploy a compositional semantics for a phrase like *tall man*? Central to most accounts is the idea that gradable adjectives, like *tall*, draw on orderings of objects for their interpretation. For example, by one account, gradable adjectives like *tall* represent functions that map objects in the world onto scales of degrees (see Cresswell, 1976; Kennedy, 1999). *Tall* maps objects onto a scale that represents degrees of height, *big* onto a scale of size, and *smart* onto a scale of intelligence. This mapping results in a mental ordering of objects that allows us to decide who is bigger, taller, or more intelligent than whom. For example, if John's degree of height is greater than Pete's, then John is taller than Pete. However, if John's degree of height is less than Pete's, John is shorter than him. Thus, adjectives create orderings of things according to the dimensions they specify (e.g., height, weight, intelligence, etc.).

For the interpretation of adjectives with comparative morphology (like *taller* and *shorter*), knowledge of object orderings alone is sufficient. The child does not require a kind-based subsective semantics for such forms because what counts as *taller* or *shorter* does not shift according to the kind of thing referred to (e.g., basketball players vs. jockeys) but merely according to the sizes of the two objects being considered. For example, if presented with two men measuring over 7 ft and asked to select ''The taller basketball player'' or even ''The tall basketball player,'' we will be forced to choose the one with the greater height, despite the fact that both are tall, even relative to other basketball players (the definite determiner forces us to make this choice in the latter example). In these cases, the standard of comparison is the height of the shorter player and is not determined relative to a kind (the result is the same if we ask which is a taller man, a taller object, etc.). However, the situation changes if we are asked whether each player is a *tall basketball player*. Now, instead of one being tall, both are. This judgment requires establishing an abstract standard of comparison that is relative to a broader set of things. This standard must shift according to the kind of thing referred to and must take account of the sizes of set members. For example, to decide whether John is a tall basketball player, the child needs to know what it means to be *tall for a basketball player* and thus have knowledge about the sizes of basketball players as a class (see Klein, 1991). Extracting this information about the sizes of things within classes thus involves performing a statistical operation—that is, obtaining a value by applying a function to a set of data.

Exactly how statistical functions are deployed to determine standards of comparison is unknown. One possibility, discussed by Bierwisch (1989), is that the value of the standard is defined as a population mean: All basketball players who are taller than the average height are *tall players*, whereas those below average are *short players.* Alternatively, the value of the standard may be specified according to a proportion of a set. *Tall* may mean *taller than most*, rather than *taller than average*, and pick out the top quarter or third of objects in a set regardless of their relation to the average height of the set. In this article, we do not take a stance on the particular statistic that children (or adults) might deploy or whether either of these alternatives is adequate. Regardless of the particular statistic that is used to determine the value of the standard, the child must know that a standard exists and have the ability to compute its value for each noun that he or she encounters. What is important is that the compositional semantics requires the child to employ data about the statistics of sets in the world and to use these statistics in identifying and referring to things as *tall* and *short*.

From previous studies, we know that children have some knowledge regarding the typical sizes of things. For example, children between 2 and 4 years of age are familiar with the typical sizes of things like buttons, mittens, plates, and shoes and are able to judge when things like mittens are big *for a mitten* (Ebeling & Gelman, 1988; Sera, Troyer, & Smith, 1988). Children are also aware that something which is small for a person, like a shot glass, may count as big for a doll (Ebeling & Gelman, 1994; Gelman & Ebeling, 1989; see also Carey & Potter, reported in Carey, 1978). These studies indicate that children know when to apply gradable adjectives like *big* and *small* to members of particular classes of familiar things. However, they do not indicate how children know this and whether they are using the compositional semantics of gradable adjectives and abstract knowledge about comparison classes to do so. Do 3- and 4-year-old children understand the compositional semantics of adjective–noun combinations like *big rabbit* and *tall man*? How does their knowledge of real world entities and their typical sizes interact with the semantics of gradable adjectives like *big*, *small*, *tall*, and *short*?

What Do Children Actually Know About Standards of Comparison?

Although knowing what counts as a *big button* could well involve knowledge of compositionality and statistics, children might initially use simpler strategies. For example, children might learn, through ostension, which mitten sizes are typically called *big* or *small*. Alternatively, they might represent a mental ordering of objects (recognizing their relative sizes) but not compute a statistical standard, instead labeling things as big or small as a function of their similarity to the biggest or tallest exemplars that they have encountered (see Clark, 1970). As noted by Smith, Rattermann, and Sera (1988), several previous studies have suggested that children might initially interpret dimensional terms like *big* and *tall* categorically, unlike adults, whereas others have argued that very young children deploy a relational interpretation (see Clark, 1970; Ehri, 1976; Gitterman & Johnston, 1983; Nelson & Benedict, 1974; Sera & Smith, 1987).

Critically, we know of no previous study that has tested what defines a comparison class for children (e.g., the class of things considered when evaluating whether something is *tall*). For adults, comparison classes are defined at least in part by nouns (e.g., a *tall tree* is tall relative to all things named by the noun tree). However, previous studies leave open the relative role of kind information in children's interpretation of adjectives, and whether other contextual factors—like contiguity, perceptual similarity, or explicit mention of class members—might play a greater role in grouping objects for comparison. For example, no previous study has investigated whether applying the same name (e.g., *blicket*) to an array of novel and physically distinct objects is sufficient for grouping these objects in a single comparison class for adjective interpretation (though see Waxman & Markow, 1995, for evidence that labels are sufficient for considering dissimilar objects as members of a single kind; also, see Ryalls, 2000, involving explicit feedback in novel adjective acquisition).

Perhaps the best existing evidence that children can interpret adjectives according to comparison classes comes from a study by Smith and colleagues on children's interpretation of *high* and *low* (Smith, Cooney, & McCord, 1986; see also Hamburger & Crain, 1984). Though other studies have tested whether children can identify big or tall items in a single novel array (e.g., Syrett, Kennedy, Bradley, & Lidz, 2006), Smith et al.'s study is alone in manipulating the distribution of object sizes (or heights, in this case) across conditions, a manipulation that is critical for assessing whether children take properties of whole sets into consideration when applying gradable adjectives. Smith et al. found that 5-year-olds but not 4-year-olds shifted their judgments of what heights counted as *high* based on the range of heights they saw presented in the study, suggesting that the older group computed a context-sensitive standard. However, words like *high* and *low* are perhaps not ideal cases for evaluating this possibility because they have both subsective meanings (e.g., high for a shelf) and intersective meanings (e.g., high off the ground, independent of kind). Although there is evidence that children can distinguish intersective adjectives from subsective adjectives (Syrett et al., 2006), it remains open whether they have a preference for one or another interpretation when both are available. Thus, it is possible that 4-year-olds understand how statistical properties of comparison classes shift standards for the interpretation of adjectives but that they, unlike older children, favor an intersective interpretation of words like *high* and *low.*

To summarize, interpreting subsective gradable adjectives like *tall* and *short* requires not only knowledge of the compositional semantics of adjective–noun combinations but also knowledge of how adjectives specify a standard of comparison based on knowledge about the size distributions of sets in the world. Currently, it is unclear when children are first able to integrate statistical properties of sets to compute standards of comparison for adjectives like *big* and *tall.* Also, there is little evidence to suggest a deep understanding of the compositionality of adjective–noun pairs. Existing studies have focused on adjective–noun pairs that children may have learned through ostension, or on comparative forms of adjectives that do not require computing standards, or they have tested adjectives that can be interpreted without appeal to the compositional semantics of adjective–noun combinations (e.g., *high*, *low*). Thus, no previous study has examined how children interpret adjectives according to statistically specified comparison classes and whether they actually use linguistic kind information to specify sets and compute standards of comparison.

This study investigated how young children deploy the compositional semantics of adjective–noun combinations in the context of novel nouns.

Doing this allowed us to address three issues. First, we investigated whether children are sensitive to the statistical properties of sets when interpreting adjective–noun combinations. By testing children's understanding of complex concepts like *tall pimwit* and *short pimwit*, we explored whether they compute standards for novel sets of objects, ensuring that previous ostensive learning could not account for their behavior. Second, we tested children's sensitivity to variation in the statistical properties of object arrays when computing a standard. Third, we explored how children restrict comparison classes. For expressions like *tall pimwit*, do children use linguistic kind information to restrict the comparison class for *tall* or do they rely initially on a nonlinguistic contextual grouping based on perceptual similarity or spatiotemporal contiguity?

Experiment 1

The first experiment explored 4-year-olds' application of the adjectives *tall* and *short* for an array of novel objects. Of interest was whether children would spontaneously consider the set as a comparison class and apply adjectives in a systematic fashion, indicative of using a standard of comparison (e.g., mean height). This condition also established a baseline for assessing children's ability to shift the value or *standard* that defines *tall* and *short* (in Experiment 2) and their ability to calculate this value based on specific kinds of objects, as specified by nouns (Experiments 3 and 4).

Method

Participants. Participants were sixteen 4-year-old children (M = 4;6, range = 4;0 – 4;11; 5 boys, 11 girls). In this experiment and those reported below, children were primarily from middle-class Caucasian families. Children were tested either in a child care center or in the laboratory. All children were native speakers of English.

Procedures and stimuli. Each testing session comprised two trials: one "short" trial and one "tall" trial. On each trial, children were asked to examine a row of pseudorandomly ordered novel objects and decide which were either tall (on tall trials) or short (on short trials). To indicate their choice, they were asked to move the relevant objects into a red plastic circle. For both types of trial, the objects were as follows: nine cylinders, ¼ in. in diameter, increasing in height from 1 to 9 in. in 1 in. intervals. The objects were painted pink and had a flat hairdo (made of foam) and a face (including googly eyes).

To begin, each child was given a categorization pretest. Sets of toy oranges, bananas, and strawberries were given to the child who was then asked to place the members of each set, one set at a time, into a red plastic circle 6 in. in diameter; for example, "Can you put the bananas into the red circle?" Following each set, the circle was emptied and the children were asked to place the next set in the circle. For each trial, all fruit types were available to the child, thus requiring them to select only the relevant kind of fruit and place it in the circle. Once children had successfully placed each set in the circle one time (with three consecutive successes), they moved to the experimental trials.

On the experimental trials, the experimenter stood the test objects in a row in one of two pseudorandom orders. The heights of objects in inches were as follows: Order 1: 9, 3, 5, 4, 8, 6, 2, 7, 1; and Order 2: 7, 9, 2, 8, 4, 1, 6, 3, 5. Objects were arranged pseudorandomly, rather than in series, to render the relations between objects less transparent and to avoid the clustering of distractor objects in later experiments. Having arranged the objects, the experimenter then said "Look! These are some pimwits. Have you ever seen any pimwits before? No? Well, can you do me a favor? Can you touch all of the pimwits like this?" The experimenter then touched each pimwit and prompted the child to do the same. This ensured that children were aware of all objects in the array. Next, the experimenter asked the child to find either the tall or short pimwits and place them in the red circle: "Can you look at all of the pimwits and find the *tall* pimwits and put the *tall* pimwits in the red circle?" or "Can you look at all of the pimwits and find the *short* pimwits and put the *short* pimwits in the red circle?" Half of the children were given the short trial first and half were given the tall trial first. Objects were returned to their original locations in the row of pimwits between *tall* and *short* trials. Thus, children were free to label an object as both *tall* and *short* (though only 2 children out of 80 reported in this article ever did so).

Following the tall–short judgments, each child was asked to perform "taller" and "shorter" judgments to assess their knowledge of comparative morphology. For taller judgments, the experimenter presented the child with the 1-in. object and the 3-in. object and asked "Which pimwit is taller?" For the shorter judgments, the child was shown the 9-in. object and the 7-in. object and asked "Which pimwit is shorter?" Thus, taller judgments were always made on objects that were likely to be labeled as *short*, and shorter judgments were made on objects that were likely to be labeled as *tall*. Children who received tall before short in the tall–short task were given the taller judgment before the shorter judgment, and the order was reversed for children who received short trials first.

Results and Discussion

Tall and short judgment. Responses for the tall judgments were coded as either tall or not tall. Responses for the short judgments were coded as either short or not short. The dependent variable for tall was the average minimum height children categorized as tall. For short, the dependent variable was the average maximum height children categorized as short. These dependent variables were used rather than mean height of the selected objects because we were interested in identifying the standard or ''cutoff'' for each adjective.

Children's average minimum selection for *tall* was 7.19 in., whereas the average maximum for *short* was 3.19 in. (see Figures 3–5). Data were submitted to a repeated measures analysis of variance (ANOVA) with adjective order (*tall* first vs. *short* first) and object order (Order 1 vs. Order 2) as between-subject variables and adjective as a within-subject variable (tall vs. short). There was a significant difference between the average minimum height of objects called *tall* and the average maximum height of objects called *short*, $F(1, 12) = 69.0$, $p < .001$, but no effect of order or interaction between adjective and order. Thus, children clearly distinguished between *tall* and *short* for a novel array of objects.

Children showed an intriguing difference between their understanding of *tall* and *short*. Though all children called the tallest object *tall*, only 11 of 16 called the shortest object short. The average maximum height of objects called *tall* was 9 in., whereas the average minimum for short was 1.56 in. For children who did not pick the shortest object as short, the average minimum was 2.8 in. ($N = 5$, range = 2–4 in.). This difference between tall and short resembles an asymmetry found in children's interpretation of *high* and *low* reported by Smith et al. (1986). Children almost always agreed that the highest heights in their experiment were *high* but agreed less frequently that the lowest values were low. These results and previous ones (e.g., Donaldson & Wales, 1970; Ehri, 1976; Ryalls, 2000; Townsend, 1976) suggest that children generally master ''positive'' terms like *tall* earlier than ''negative'' terms like *short*.

On several occasions during this experiment and in those reported later, children insisted that the shortest objects in a set were not *short* but instead were *small* or *little.* This can be interpreted in one of two ways. On one hand, given the observation that children acquire the terms *big* and *small* before *tall* and *short* (Clark, 1973; Maratsos, 1973), they may be reluctant to apply *short* and *small* to the same objects. Mutual exclusivity (Markman, 1988) may lead them to assign middle values to *short*, with no other meaning available (presumably the greater frequency and earlier acquisition of *big* and *tall* could explain why children in this study do not show a similar effect for *tall*). However, a second possibility is that children, like adults, place geometric restrictions on the kinds of things that can be placed on the *tall–short* continuum (see Lang, 1989). The shortest objects may have been called *small* not because children failed to understand *short* but because, as with pebbles or balls, very short pimwits did not feature a sufficiently large ratio between their height and width to merit the term.

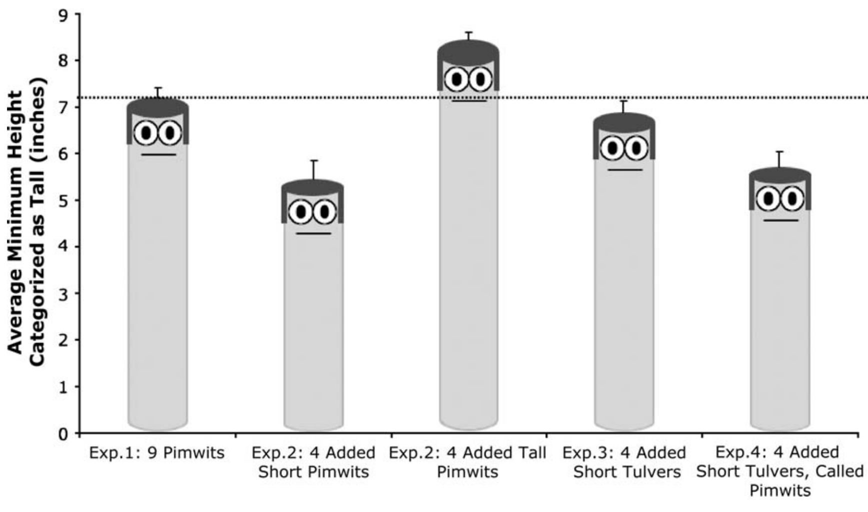

Figure 3. The average minimum height categorized as ''tall'' for Experiments 1–4.

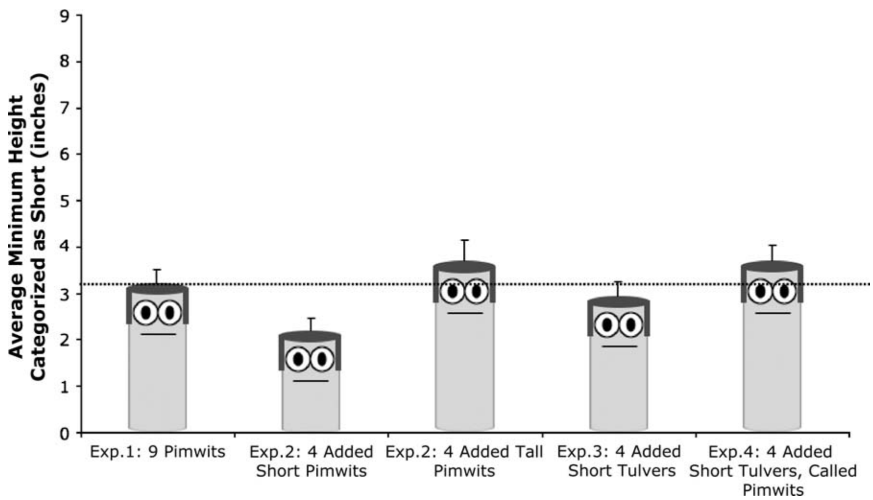

Figure 4. The average maximum height categorized as "short" for Experiments 1–4.

Comparatives: Taller and shorter judgments. Judgments for the comparative forms *taller* and *shorter* were available for 15 of the 16 children. All 15 answered correctly to both the *taller* and the *shorter* questions. Also, all the children who failed to call the shortest object *short* in the first part of the experiment nonetheless had no difficulty interpreting *shorter*, even in the context of tall objects. Thus, although we found an asymmetry between *tall* and *short* in the first part of the experiment, we did not replicate previous reports (e.g., Ryalls, 2000) that children of this age have difficulty interpreting *shorter* relative to *taller*. Further, we failed to find this effect even though children were asked to pick the shorter of two tall objects or the taller of two short ones. In Experiments 2–4, all children (N = 64) responded correctly to *taller* and *shorter* judgments, and thus results for those conditions will not be reported further.

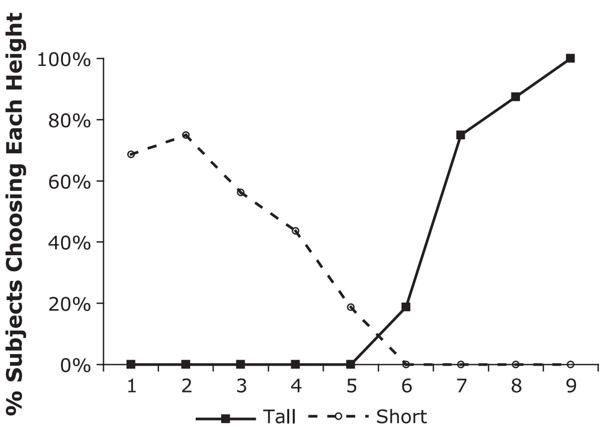

Figure 5. Percentage of children who judged each height of object to be tall or short in Experiment 1.

Summary. To conclude, we found that 4-year-olds were able to systematically select tall and short items from an array of novel objects. Further, we found an asymmetry between *tall* and *short* judgments but no corresponding asymmetry for *taller* and *shorter* judgments. This result resembles the asymmetry found by Smith et al. (1986) between *high* and *low*.

Experiment 2

The second experiment assessed children's sensitivity to the statistical properties of object arrays when applying the adjectives *tall* and *short*. Previous studies of children of this age have demonstrated that their interpretation of adjectives is flexible and influenced by the objects that are available in a context (e.g., Sera & Smith, 1987; Syrett et al., 2006). However, these studies did not address the role of kind information or require that statistical information be used to calculate the standard of comparison because they tested children with comparisons among object pairs (akin to studies that examine the interpretation of comparative forms like *taller* and *shorter*).

To assess whether 4-year-old children can access statistical information regarding the sizes of set members to calculate a standard, we tested children in two conditions using modified versions of the object arrays used in Experiment 1. In the short distractor condition, four distractor objects were added to the original array of nine objects. The distractors were identical to those in the original array but were relatively short: 0.5, 1, 1.5, and 2 in. Adding these items reduced the mean height of objects in the array from 5 to 3.85 in. In the tall distractor condition, the four additional objects measured 8, 8.5, 9, and 9.5 in., thereby increasing the average height of objects from 5 to 6.15 in. We predicted that if children are sensitive to the statistical properties of the arrays, they should shift their tall and short judgments accordingly. When presented the baseline set with shorter distractors, they should include more of the original nine objects as tall and fewer of the original nine as short. For taller distractors, they should include fewer of the original objects as tall and more as short.

Method

Participants. Participants were thirty-two 4-year-olds (M = 4;6, range = 3;11 – 4;11.15; 18 boys, 14 girls), who had not participated in Experiment 1. Children were tested either in a child care center or in the laboratory. All children were native speakers of English. Half were assigned to Condition 1 (shorter distractors; M age = 4;5) and half to Condition 2 (taller distractors; M age = 4;6).

Procedures and stimuli. Procedures for Experiment 2 were identical to those used in Experiment 1. However, in addition to the nine original objects, there were four distractor objects. For Condition 1, four short distractor objects were added, measuring: 0.5, 1, 1.5, and 2 in. These distractor objects were mixed pseudorandomly among the other nine objects to produce two different array orders. Objects were ordered as follows in Orders 1 and 2 of Condition 1 (distractors are indicated with < >): Order 1: 9, 3, <0.5>, 5, 4, <1>, 8, 6, <1.5>, 2, 7, <2>, 1; and Order 2: 7, 9, <0.5>, 2, 8, <1>, 4, 1, <1.5>, 6, 3, <2>, 5.

For Condition 2, four tall distractor objects were added, measuring: 8, 8.5, 9, and 9.5 in. The objects were ordered as follows for Orders 1 and 2 of Condition 2: Order 1: 9, 3, <8>, 5, 4, <8.5>, 8, 6, <9>, 2, 7, <9.5>, 1; and Order 2: 7, 9, <8>, 2, 8, <8.5>, 4, 1, <9>, 5, 6, 3, <9.5>, 5. In familiarization, children were asked to touch each of the pimwits on its head, and thus they touched the distractor objects in addition to the target objects.

Results and Discussion

To assess the effect of adding shorter items versus taller items to the novel object arrays, we compared *tall* and *short* judgments from Conditions 1 and 2. First, for *tall* judgments, data were submitted to an ANOVA with adjective order (tall first vs. short first), object order (Order 1 vs. Order 2), and condition (short distractors vs. tall distractors) as between-subject variables. The dependent variable was the average minimum height of objects called *tall*. Children picked a greater minimum height on average for *tall* judgments in the tall distractor condition (M = 8.44 in.) compared to the short distractor condition (M = 5.44), $F(1, 24) = 44.88$, $p < .001$ (see Figures 3, 4, 6, and 7). There were no main effects or interactions due to order of adjectives or object order. Second, a parallel ANOVA found that children picked a greater maximum height on average for *short* judgments in the tall distractor condition (M = 3.69) compared to the short distractor condition (M = 2.19), $F(1, 24) = 8.08$, $p < .01$. Again, there were no significant main effects or interactions due to order of adjectives or object order. These results indicate that differences between the two arrays had significant effects on children's interpretation of both *tall* and *short.*

We next compared each condition of Experiment 2 to the baseline condition of Experiment 1. First, consider the short distractor condition. Children picked a greater minimum height on average for *tall* judgments in Experiment 1 (M = 7.19) compared to judgments in Condition 1 (M = 5.44), $F(1, 24) = 12.92$, $p < .001$. There were no main effects or interactions due to order of adjectives or object order. Second, children picked a greater maximum height on average for *short* judgments in Experiment 1 (M = 3.19) compared to Condition 1 (M = 2.19), $F(1, 24) = 4.68$, $p < .05$. Again, there were no main effects or interactions due to order of adjectives or object order. Thus, when four short objects of the same kind were added to the original array of Experiment 1, the standards for both *tall* and *short* were lower relative to Experiment 1.

For the tall distractor condition, objects that children categorized as *tall* had a greater minimum height on average (M = 8.44 in.) relative to Experiment 1

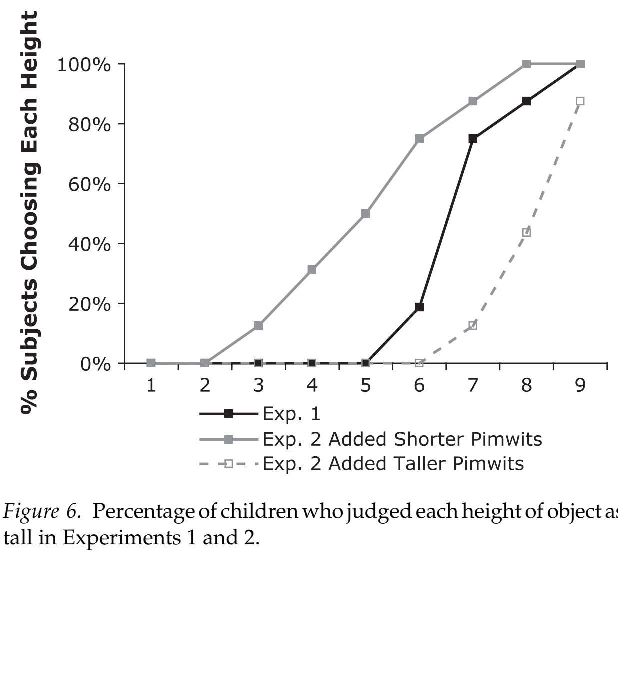

Figure 6. Percentage of children who judged each height of object as tall in Experiments 1 and 2.

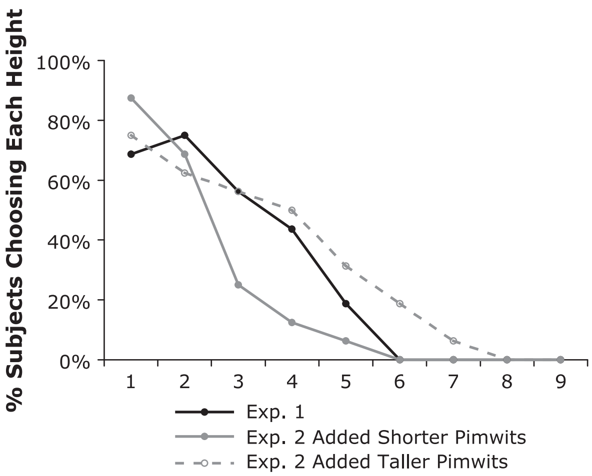

Figure 7. Percentage of children who judged each height of object as short in Experiments 1 and 2.

(7.19 in.), $F(1, 24) = 17.14$, $p < .001$. There were no main effects or interactions due to order of adjectives or object order. However, their judgments for *short* did not differ significantly between Experiment 1 (3.19 in.) and Condition 2 of Experiment 2 ($M = 3.69$ in.), $F(1, 24) = 0.81$, $p > .05$, though there was a trend in the predicted direction. Thus, although the addition of tall distractors had a significant effect on children's criteria for the application of *tall*, it did not in this case appear to affect their application of *short*, at least relative to Experiment 1. One observation relevant to interpreting children's application of *short* is the fact that several children again failed to call the shortest objects *short* in both conditions of Experiment 2 (two in each condition). This suggests that children may still be acquiring the semantics of *short* and may not yet be in a position to systematically shift its application as a function of height distributions in a set. We return to this question later in the article.

To summarize, two manipulations generated evidence that 4-year-olds can shift their standard of comparison appropriately for *tall* when the statistical properties of the object array were modified. Children also exhibited signs of understanding how differences in height distributions shift the application of *short*, though this effect was less robust.

Experiment 3

Experiments 1 and 2 indicate that 4-year-old children are able to extract statistical information from a novel array of objects and compute an implicit standard of comparison for the application of adjectives like *tall* and *short*. This capacity is sensitive to small shifts in the statistical properties of object arrays, with significant differences being incurred by the addition of a few additional objects.

In order to compute a standard for interpreting comparative adjectives, children must presumably first determine what the class of objects is that is relevant to comparison. For adults, this is determined in large part on the basis of kind information for expressions like "find the tall pimwits." In such cases, things are deemed tall or short relative to a kind of thing. However, no previous study has examined how children segregate novel objects into discrete comparison classes, and if this is based on kind information. In other constructions, the standard of comparison is determined contextually. For example, the expression "that watch is expensive" may be true in a context where the watch is only $5 but is found at a garage sale where all other objects are below $1. This differs from "that is an expensive watch," which almost always means "expensive for a watch." Children may not realize this distinction early in acquisition and may be sensitive to context (e.g., the properties of adjacent objects) when interpreting adjectives in both types of construction.

Experiment 3 investigated 4-year-olds' sensitivity to kind information in establishing a comparison class and whether proximity with different-kind distractors interferes with the use of kind information. To do this, the original nine objects of Experiment 1 were presented to children together with four additional objects that were identical to the four short distractors added in Condition 1 of Experiment 2, except that they were of a different kind (i.e., different color, surface features, name). Thus, all that differed from the short distractor condition of Experiment 2 was the kind status of the additional objects. Of interest was whether these different-kind objects would interfere with judgments for the other nine objects (resulting in judgments like those from Experiment 2) or whether children would use kind information to exclude these items from the comparison class (resulting in judgments like those from Experiment 1).

Method

Participants. Participants were 16 English-speaking children not tested in Experiments 1 and 2 (M = 4;5, range = 4;0 – 4;11; 13 boys, 3 girls). Children were tested either in a child care center or in the laboratory.

Procedures and stimuli. Procedures for Experiment 3 were identical to those used in the short distractor condition of Experiment 2, except that the four distractor objects were of a different kind. Distractors were cylinders like those in Experiment 2 but were painted silver with black dots resembling rivets and had flat hexagon shapes on either end (resulting in dumbbell-like objects). Also, they were called *tulvers* rather than *pimwits.* The distractor objects were mixed pseudorandomly among the other nine objects to produce two different array orders (identical to the orders used in the short distractor condition of Experiment 2). In familiarization, children were asked to touch the pimwits and were also told that the distractor items were tulvers and to touch these as well.

Results and Discussion

The average minimum height of objects called *tall* for the nine target items was 6.89 in., whereas the average maximum height of objects called *short* was 2.94 in. (see Figures 3 and 4). Data were submitted to a repeated measures ANOVA parallel to that in Experiment 1. There was a significant difference between the average minimum height of objects called *tall* and the average maximum height of objects called *short*, $F(1, 12) = 73.05$, $p < .001$. There was no main effect of order or interaction between adjective and order. As in previous conditions, 4-year-olds had an imperfect understanding of the adjective *short*. Though all children called the tallest object tall, 7 out of 16 children did not call the shortest object short, including one who said that no objects were short.

We compared the results of this experiment with those from the short distractor condition of Experiment 2, in which short distractors received the same name as targets. Judgments made in the context of short same-kind distractors (Experiment 2) were significantly lower than those made in the context of short different-kind distractors (Experiment 3) for tall trials (5.44 in. vs. 6.89 in.), $F(1, 24) = 8.87$, $p < .01$, and for short trials (2.19 in. vs. 2.94 in.), $F(1, 24) = 3.0$, $p < .05$, one tailed. However, a comparison of Experiments 1 and 3 revealed no differences between the baseline condition (Experiment 1) and the different-kind condition (Experiment 3) for either tall trials (7.19 in. vs. 6.89 in.) or short trials (3.19 in. vs. 2.94 in.). This suggests that children successfully attended to kind information in order to establish the standard of comparison for the target comparison class. Added objects only affected *tall* and *short* judgments for the target set when they were of the same kind.

Experiment 4

Experiment 3 suggests that 4-year-old children are sensitive to kind information when assessing the statistical properties of comparison classes. However, the experiment leaves open the question of whether children's creation of distinct comparison classes for the two types of objects (*pimwits* and *tulvers*) was based on the objects having distinct physical properties or distinct labels. Only the second possibility would provide evidence for a compositional analysis in which meaning of the adjective depended on the meaning of the noun per se. Thus, it is unclear whether children's comparison classes are based on *linguistic* criteria (i.e., kind categories delineated by language) or on nonlinguistic conceptual or perceptual categories. To address this question, we tested children using the objects of Experiment 3 (i.e., pimwits and distractor tulver objects) but called all the objects *pimwits*. If children include objects in comparison classes on the basis of their names, then they should shift their judgments relative to the baseline and show judgments similar to those found in Experiment 2, in which distractors were physically identical to target items. However, if their comparison classes are based on perceptual or basic-level conceptual categories, then they should fail to shift their judgments relative to baseline and show judgments similar to Experiment 3.

Method

Participants. Participants were 16 English-speaking children not tested in Experiments 1–3 (M = 4;6, range = 4;0 – 4;11; 8 boys, 8 girls). Children were tested either in a child care center or in the laboratory.

Procedures and stimuli. The target and distractor objects were identical to those used in Experiment 3, and as before, the distractors were mixed pseudorandomly among the nine targets to produce two different array orders (identical to the orders used in the short distractor condition of Experiment 2). However, in familiarization, children were given additional training to be certain that they understood all objects were *pimwits*. They were asked not only to touch the *pimwits* but also told that some *pimwits* are pink and some are gray and that some have eyes and that some do not have eyes. Children were asked to find a *pimwit* with eyes and one without eyes or to find a gray *pimwit* and a pink *pimwit*. In this way, we were sure that children understood that all objects were *pimwits*.

Results and Discussion

The average minimum height of objects called *tall* for the nine target items was 5.69 in., whereas the average maximum height of objects called *short* was 3.69 in. (see Figures 3 and 4). Data were submitted to a repeated measures ANOVA parallel to that in

Experiment 1. There was a significant difference between the average minimum height of objects called *tall* and the average maximum height of objects called *short*, $F(1, 12) = 14.63$, $p < .001$. There was no effect of order or interaction between adjective and order.

A comparison of Experiments 3 and 4 revealed that when the perceptually different distractor items were given the same name as target items, the average height of objects called *tall* was significantly lower than it was when distractors were given a different name (5.69 in. vs. 6.89 in.), $F(1, 24) = 8.14$, $p < .005$. Further, *tall* judgments for target items did not differ from those in Experiment 2, where distractor objects were both physically similar to target items and were given the same name, $F(1, 24) = 0.22$, $p > .6$.

However, as in the tall distractor condition of Experiment 2, short judgments did not shift in the expected way. First, there was no difference between *short* judgments here relative to those in Experiment 3 (M = 3.69 and 2.94 in., for Experiments 3 and 4, respectively), $F(1, 24) = 2.27$, $p > .1$. Second, there was a significant difference between *short* judgments in this experiment and Experiment 2. The maximum height called *short* was significantly higher when distractors with the same name were perceptually different from targets than when they were perceptually similar, $F(1, 24) = 9.00$, $p < .01$.

We see two plausible accounts of why children's judgments for *short* do not shift as they do for *tall*. First, it is possible that many children have not yet acquired the meaning of *short*. In Experiment 4, 5 of 16 children did not call the shortest target item *short*. Among these children, the average maximum of objects called *short* was 4.2 in. compared to 3.45 in. for the remaining children. In addition, 6 of 16 did not call the even shorter distractor items short. Further, as noted earlier, some children insisted that the shortest objects were not short because they were *small*, suggesting that the two may initially have mutually exclusive meanings. Another possible explanation is that children of this age do understand *short* but fail to assign the shortest pimwits to the tall – short scale because the objects lack a sufficiently large height: width ratio. By this account, children prefer to call the shortest objects *small* as these objects fit onto the *big – small* scale but not the *tall – short* scale. In either case, data from *tall* judgments may provide a more reliable indicator of the effect of linguistic information on the creation of comparison classes because the children show greater mastery of this term.

In summary, in Experiment 4, when distractor items were physically different from targets but were given the same name, children's *tall* judgments shifted to the same extent as in the short distractor condition of Experiment 2. For *tall* at least, the kind term was sufficient for including both targets and distractors in a single comparison class.

General Discussion

By at least 4 years of age, children learning English are able to derive composed meanings from novel adjective – noun combinations and can rapidly integrate subtle changes in the statistical properties of object arrays to shift their standard for applying adjectives. At least for *tall*, this sensitivity to statistical information is guided mainly by linguistic cues to categorization: Objects that have different features but the same name are considered when applying an adjective – noun pair, suggesting that, for 4-year-olds, linguistic information is sufficient for creating comparison classes.

Integrating Semantics and World Knowledge

Knowing the meaning of the adjective *tall* does not alone indicate how the word should be applied to objects in the world. On its own, the adjective meaning supports inferences but not the application of the term. For example, knowing that a pimwit is tall tells us that pimwits are the types of things that have vertical extension. We can also infer that if a pimwit is tall, then it is likely taller than most other objects of its kind, independent of its absolute height. Further, we can infer that a *tall pimwit* is not short for a pimwit. However, without knowledge about the range of heights that are typical for pimwits, we cannot judge whether a given pimwit is tall or short because things can only be labeled as such relative to other individuals. For familiar kinds of things, semantic, inferential, knowledge is supplemented by world knowledge regarding the typical sizes of kind members. This type of world knowledge allows us to apply *tall* to people, trees, and buildings. Thus, to apply adjectives like *tall*, children need to acquire not only the compositional semantics of gradable adjectives but also how this semantics relates to the typical sizes of kinds of things in the world.

How is world knowledge combined with the compositional semantics of adjectives? As noted earlier, many accounts propose that gradable adjectives represent functions that result in the ordering of objects according to a scale (e.g., of height, width). This scale may be composed of degrees (e.g., degrees of height; Cresswell, 1976; Kennedy, 1999) or may order objects without appeal to degree semantics

(Klein, 1991). For each scale, adjectives like *tall* and *short* specify positions in the ordering. Thus, no appeal to knowledge of things in the world is required to grasp the semantics of the adjective because relations between adjectives and degrees can be stated independent of actual measures.

To use this knowledge for identifying *tall* and *short* things, the child must amass a database of typical object sizes (e.g., the heights of trees, boys) and divide each class into *tall* and *short* subclasses. This requires setting a value, or standard, against which set members can be compared and classified as tall or short. For example, children may assume that for all gradable adjectives, the standard of comparison is defined by the arithmetic mean, such that *tall tree* applies to all trees that are taller than average. Alternatively, they may assume that *tall* refers to all degrees above the third quartile of values (i.e., that it means *taller than most objects*). Whichever standard is used, the child must apply it to data for particular kinds of things in order to determine its value. For example, assuming that the standard is defined as an arithmetic mean, the child would need to collect information about the average heights of trees to set the value of the standard for the kind TREE. Having determined this value, the child would then be ready to apply the adjective *tall* to trees in the world.

The question of how semantic structures and world knowledge are related leaves open the processes by which they *become related* in development. Do they emerge independently or does one depend on the other in acquisition? One possibility, which we label the *innate standards* hypothesis, is that the semantic relation between adjectives and standards is not learned and that upon realizing that a novel word is gradable (perhaps via syntactic cues like comparative morphology), children assume that it has an opposite and that these opposites exhaustively divide sets in two, or alternatively, result in a three-way categorization of objects (e.g., *tall*, *short*, and neither *tall* nor *short*). Given this assumption, children might then make use of whatever domain general tools they have available for assigning absolute values (i.e., object sizes) to this division (e.g., using prototype structures that represent mean size). According to this view, their acquisition of statistical information relevant to defining standards of comparison could take place independent of their acquisition of semantic knowledge, and the two could be paired up when sufficient statistical information is available. Alternatively, by the *constructed standards* hypothesis, children might infer how adjectives name subsets of object orderings (i.e., what in fact defines the statistical standard) by observing how adjectives are used by others to describe particular things in the world (e.g., that men 5 ft 11 in. and taller are called *tall men*, and 5 ft 8 in. and shorter are called *short men*). By learning such orderings via ostension for a subset of nouns, children might then abstract the notion of a standard and apply the notion to novel objects in absence of ostensive evidence. Thus, a bottom-up ordering of objects might be used to infer the existence of a standard and what defines it.

The data presented here address only certain aspects of this learnability question. The data demonstrate that children do not need to learn all standards of comparison by ostension (e.g., by hearing certain things called *tall* and other things called *short*). Instead, they can compute a standard based on the properties of things being referred to. Thus, by at least 4 years old, and possibly earlier, children have an internalized set of criteria for establishing standards of comparison for known adjectives, which can be applied in a general fashion to new kinds of things. To explore the question of how this ability develops, two additional paths of investigation suggest themselves. First, on the constructed standards account (where the notion of a standard and its definition are abstracted from parental language use), the way in which standards are specified for adjectives should be subject to substantial variation, whether between children or cross-linguistically. Because the standard is determined by the input on this account, children's conception of the possible divisions of sets by adjectives should be malleable. Therefore, if told from the beginning that the tallest three fourths of trees are tall trees (e.g.), children should happily abstract a standard that treats the tallest three fourth of pimwits as *tall pimwits.* However, by the innate standards view, where semantic representations emerge independent of statistical knowledge of particular kinds, little or no variation is expected between children or across cultures, and children's assumptions about adjective meanings should be relatively impervious to training. Even if explicitly told that the tallest three fourth of various sets are *tall*, very young children should still retain a semantics that divides sets as children do in the current study (e.g., where roughly the top one third are called tall).

Second, if children used world knowledge regarding the typical sizes of things to bootstrap their semantic representations for gradable adjectives, we would expect them to acquire knowledge about the typical sizes of things sometime before they understand the compositional semantics of adjective – noun combinations. Based on this, it is imperative to compare the behavior of 4-year-olds with younger children who are just beginning to acquire gradable adjectives. As noted in the Introduction, previous studies have documented that even 2-year-olds have some knowledge of the typical sizes of household objects, suggesting that this type of world knowledge is acquired early in development (Ebeling & Gelman, 1988; Gelman & Ebeling, 1989; Sera et al., 1988; see also Carey & Potter, reported in Carey, 1978). Also, Smith et al. (1986) present evidence that for 5-year-olds but not 4-year-olds, the reference points for *high* and *low* are affected by changes in the range of heights presented to children, suggesting that the younger children were not computing a context-sensitive standard under these conditions. However, as noted earlier, it is unclear whether adjectives like *high* interact with kind information in the same way as *tall* to derive composed meanings (i.e., it is unclear that children *should* acquire a standard for interpreting *high bunny* in the same way they might for *tall man* because being tall is a property that gets its meaning from the relation between objects, whereas being high is not a relation that exists between objects of a kind but between objects and the ground).

Future studies should examine whether there is a stage at which very young children have knowledge of the typical sizes of things in absence of a capacity to deploy a standard of comparison for novel sets. Here, we have shown that 4-year-olds have combined these two forms of knowledge, and we have provided a paradigm which can be extended to younger children, in order to explore the way in which these representations are acquired.

Falling Short of Short

Children's mastery of *short* may not be complete by 4 years of age. In each of the experiments reported here, children exhibited an asymmetry between their judgments for *tall* and *short*. In all, 21 of the 80 children tested (26.3%), failed to call the shortest object *short*, despite always calling the tallest object *tall* and despite correctly interpreting the comparative forms of both *tall* and *short* (i.e., *taller*, *shorter*). This finding is consistent with a large body of literature that has examined children's acquisition of positive and negative polar terms, including *big* versus *little*, *tall* versus *short*, and *more* versus *less*. In general, when children exhibit differences in their ability to understand one of a pair of adjectives, it is the negative polar item that exhibits a delay (e.g., Donaldson & Wales, 1970; Ehri, 1976; Townsend, 1976; see Johnston, 1985, for a review).

However, the exact source of the asymmetry found in this study remains unclear. Children who failed to call the shortest objects *short* may have done so because they lacked an adult interpretation of the word or may have understood the word but thought that it did not apply to objects that were extremely short. As noted earlier, children insisted on several occasions that the shortest objects in a set were not *short* but instead were *small* or *little.* This could be attributed to a mutually exclusive interpretation of *short* and *small* (where small applies to the shortest objects and *short* to middling values), or alternatively, could be due to a geometric constraint. For example, some children may have refused to call very short objects *short* because the ratio between their vertical and horizontal extents was not sufficiently large. As noted by Lang (1989) geometric constraints of this kind restrict the application of adjectives, precluding the use of *tall* to name objects that typically have equal horizontal and vertical extents (e.g., like a pebble; see also Fillmore, 1977). Such a constraint would explain not only children's reluctance to call very short things *short* but also why they occasionally fail to consider very short things when calculating the standard of comparison, as in Experiments 2 and 4: Things that lack *tallness* altogether may not count as inputs to determining the average height of a set because they are not mapped to the relevant scale. To contrast this hypothesis with the possibility that *short* is mutually exclusive with *small*, future studies should test children using objects where the shortest objects have a greater height:width ratio (e.g., using either taller or skinnier objects).

Summary: Knowing What It Means to Be a Tall Pimwit

At the core of the problem of language acquisition is the question of when, and how, children begin analyzing and producing expressions using a generative syntactic system. Creative language use depends on understanding compositionality—how the meanings of complex expressions are a function of their syntax and the meanings of their constituent parts. This study suggests that by at least 4 years of age, children draw on knowledge of compositionality to interpret novel adjective – noun combinations. Without being told which pimwits are *tall* or *short*, children can apply these adjectives in a systematic fashion based on small amounts of experience. Further, the study indicates that this compositional knowledge is linked to sensitive statistical representations. Even small shifts in the average height of objects (e.g., 2 in.) are detectable by children and license shifts in how they apply adjectives to novel sets. Thus, by 4 years of age, children appear to use a compositional semantics for interpreting subsective adjective expressions like *tall pimwit*, suggesting that by at least this age, adjective – noun combinations engage a rich understanding of syntax and semantics for their interpretation.

References

Bierwisch, M. (1989). The semantics of gradation. In M. Bierwisch & E. Lang (Eds.), *Dimensional adjectives* (pp. 71 – 261). Berlin, Germany: Springer-Verlag.

Carey, S. (1978). The child as word learner. In J. Bresnan, G. Miller, & M. Halle (Eds.), *Linguistic theory and psychological reality* (pp. 264 – 293). Cambridge, MA: MIT Press.

Clark, E. V. (1973). What's in a word? On the child's acquisition of semantics in his first language. In T. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 65 – 110). New York: Academic Press.

Clark, H. H. (1970). The primitive nature of children's relational concepts. In J. R. Hayes (Ed.), *Cognition and the development of language*. New York: Wiley.

Cresswell, M. J. (1976). The semantics of degree. In B. H. Partee (Ed.), *Montague grammar* (pp. 261 – 292). New York: Academic Press.

Donaldson, M., & Wales, R. J. (1970). On the acquisition of some relational terms. In J. R. Hayes (Ed.), *Cognition and the development of language* (pp. 235 – 268). New York: Wiley.

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

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

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

Fillmore, C. (1977). Scenes-and-frame semantics. In A. Zampolli (Ed.), *Linguistic structures processing* (pp. 55 – 81). Amsterdam: North Holland.

Fodor, J. A., & Lepore, E. (2002). *The compositionality papers*. Oxford, UK: Clarendon Press.

Frege, G. (1892). On concept and object. Reprinted in B. McGuiness (Ed.), *Collected papers on mathematics, logic, and philosophy* (pp. 182 – 194). Oxford, UK: Blackwell.

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

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

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

Hall, D. G., Waxman, S. R., & Hurwitz, W. (1993). How 2- and 4-year-old children interpret adjectives and count nouns. *Child Development*, *64*, 1651 – 1664.

Hamburger, H., & Crain, S. (1984). Acquisition of cognitive compiling. *Cognition*, *17*, 85 – 136.

Johnston, J. R. (1985). Cognitive prerequisites: The evidence from children learning English. In D. Slobin (Ed.), *The crosslinguistic study of language acquisition: Vol. 2. Theoretical issues* (pp. 961 – 1004). Hillsdale, NJ: Erlbaum.

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

Kennedy, C. (1999). *Projecting the adjective: The syntax and semantics of gradability and comparison*. New York: Garland.

Klein, E. (1991). Comparatives. In A. von Stechow & D. Wunderlich (Eds.), *Semantics: An international handbook of contemporary research* (pp. 673 – 691). Berlin, Germany: Walter de Gruyter.

Klibanoff, R. S., & Waxman, S. R. (2000). Basic level object categories support the acquisition of novel adjectives: Evidence from preschool-aged children. *Child Development*, *71*, 649 – 659.

Lang, E. (1989). The semantics of dimensional designation of spatial objects. In M. Bierwisch & E. Lang (Eds.), *Dimensional adjectives: Grammatical structure and conceptual interpretation* (pp. 263 – 417). New York: Springer-Verlag.

Lieven, E., Behrens, H., Speares, J., & Tomasello, M. (2003). Early syntactic creativity: A usage based approach. *Journal of Child Language*, *30*, 333 – 370.

Maratsos, M. P. (1973). Decrease in the understanding of the word ''big'' in preschool children. *Child Development*, *44*, 747 – 752.

Markman, E. M. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, *20*, 1 – 27.

Matthei, E. M. (1982). The acquisition of prenominal modifier sequences. *Cognition*, *2*, 301 – 332.

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

Mintz, T., & Gleitman, L. (2002). Adjectives really do modify nouns: The incremental and restricted nature of early adjective acquisition. *Cognition*, *84*, 267 – 293.

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

Olguin, R., & Tomasello, M. (1993). Twenty-five-month-old children do not have a grammatical category of verb. *Cognitive Development*, *8*, 245 – 272.

Partee, B. (1995). Lexical semantics and compositionality. In D. Osherson (General Ed.) & L. Gleitman & M. Liberman (Eds.), *Invitation to cognitive science. Part I: Language* (2nd ed., pp. 311 – 360). Cambridge, MA: MIT Press.

Ryalls, B. (2000). Dimensional adjectives: Factors affecting children's ability to compare objects using novel words. *Journal of Experimental Child Psychology*, *76*, 26 – 49.

Sera, M., & Smith, L. B. (1987). Categorical and relative interpretations of ''big'' and ''little.'' *Cognitive Development*, *2*, 89 – 112.

Sera, M., Troyer, D. E., & Smith, L. B. (1988). What do two-year-olds know about the sizes of things? *Child Development*, *59*, 1489 – 1496.

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

Smith, L. B., Rattermann, M. J., & Sera, M. (1988). ''Higher'' and ''lower'': Comparative and categorical interpretations by children. *Cognitive Development*, *6*, 131 – 145.

Syrett, K., Kennedy, C., Bradley, E., & Lidz, J. (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 GALANA*. Cambridge, MA: MITWPL.

Tomasello, M. (1992). *First verbs: A case study of early grammatical development*. Cambridge, UK: Cambridge University Press.

Tomasello, M. (2000). The item based nature of children's early syntactic development. *Trends in Cognitive Sciences*, *4*, 156 – 163.

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

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

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