

This paper is to appear in *Language Learning and Development*

Title: Syntactic and lexical inference in the acquisition of novel superlatives

Journal: *Language Learning and Development*

Authors:

Alexis Wellwood

Department of Linguistics

Northwestern University

2016 Sheridan Rd

Evanston, IL 60208

Email: wellwood@northwestern.edu

Annie Gagliardi

Institute for Language, Cognition and Computation

Informatics Forum (3.29)

10 Crichton St Edinburgh, UK

EH8 9AB

Email: agaglia2@inf.ed.ac.uk

Jeffrey Lidz

Department of Linguistics

University of Maryland

1401 Marie Mount Hall

College Park

20742

Email: jlidz@umd.edu

Title: Syntactic and lexical inference in the acquisition of novel superlatives

Abstract: Acquiring the correct meanings of words expressing quantities (*seven, most*) and qualities (*red, spotty*) present a challenge to learners. Understanding how children succeed at this requires understanding not only what kinds of data are available to them, but also the biases and expectations they bring to the learning task. The results of our word-learning task with 4 year-olds indicate that a “syntactic bootstrapping” hypothesis correctly predicts a bias towards quantity-based interpretations when a novel word appears in the syntactic position of a determiner, but leaves open the explanation of a bias towards quality-based interpretations when the same word is presented in the syntactic position of an adjective. We develop four computational models that differentially encode how lexical, conceptual, and perceptual factors could generate the latter bias. Simulation results suggest it results from a combination of lexical bias and perceptual encoding.

1 Words for quantities and qualities

Learning novel words is a domain in which young children are expert: as one conservative estimate puts it, they busily acquire around 10 words per day from the time they are one year old, achieving a lexicon of approximately 12,000 words by the time they are six years old, all with nearly no effort or explicit instruction (Bloom 2000, 2002; also Anglin 1993). How do they do this? It is uncontroversial that children make use of both linguistic information (the sequences of sounds they hear) and extralinguistic information (the context of speech) when they set about learning the meaning of novel words, but the idea that some pairing of “situation and sound” is *sufficient* has been repeatedly questioned (e.g., Landau & Gleitman 1985, Waxman & Lidz 2006).

Approaching the question first requires an appreciation of the kinds of word meanings that are the target of acquisition.¹ Some words refer to object categories (*dog, mammal*) and others to event categories (*run, watch*): in acquiring such words, simply paying attention to the right aspects of the environment could in principle provide strong evidence that a novel word has a certain sort of meaning. However, this is only the very beginning of the story: many words refer to properties of objects or events (*red, fluffy, fast, suddenly*), and others refer to nothing at all (*most, any, empty*). Since any novel word could express innumerably many things, properties, or relations, understanding how children decide what a novel word means must be informed not only by a precise understanding of the kinds of data children have available to them, but also of the character of the biases and expectations they bring to the learning task.

An especially difficult problem for any view that posits a simple mapping from a portion of experience to the meaning of a novel word has been the acquisition of number words (e.g., *five, sixty-seven*). The particular challenge that words from this domain pose is that the properties number words express are quite abstract: they refer to properties of *sets* of objects (Frege 1893; Bloom & Wynn, 1997). Research on the acquisition of exact number words suggests that language itself may provide critical support for the child to map new words onto such abstract meanings. Wynn

¹Notice this already moves well beyond many other tasks that face the learner, such as parsing the speech stream into phonological and morphological units, etc.

(1992; see also Condry & Spelke 2008) found that children at the age of 2 years 6 months, who do not yet understand the relationship between the words in the count list and exact cardinalities, nevertheless understand that the number words express numerosity. This result is striking, as it takes children another full year to gain the knowledge of which exact quantities are intended (Wynn 1992, Carey 2009).

Examining the distribution of numerals in the CHILDES database of child-directed speech, Bloom and Wynn (1997) proposed that the appearance of an item in the partitive frame (e.g., as *X* in *X of the cows*) was a strong cue to number word meaning. The plausibility of this hypothesis finds support from the linguistics literature: partitivity has been called a signal to the semantic role of quantification (Jackendoff 1977). Considering a sentence like (1) with the novel word *gleeb*, it is plain to the adult speaker of English that this word cannot express anything but a numerical property of the set of cows. This intuition follows from the knowledge that grammatical substitutions of *gleeb* can only express numerical properties of sets, (1a-e).

(1) Gleeb of the cows are by the barn.

- a. * **Red** of the cows are by the barn. *COLOR
- b. * **Soft** of the cows are by the barn. *TEXTURE
- c. * **Big** of the cows are by the barn. *SIZE
- d. **Many** of the cows are by the barn. APPROX. NUMBER
- e. **Seven** of the cows are by the barn. PRECISE NUMBER

In other grammatical frames, such strong intuitions are not observed: adult English speakers allow for the novel word in (2) to express any number of properties that might be instantiated by a group of cows, an intuition that, again, likely follows from the knowledge that substitution instances of the novel word can express numerical or non-numerical properties, (2a-2e).

(2) The gleeb cows are by the barn.

- a. The **red** cows are by the barn. COLOR
- b. The **soft** cows are by the barn. TEXTURE

c.	The big cows are by the barn.	SIZE
d.	The many cows are by the barn.	APPROX. NUMBER
e.	The seven cows are by the barn.	PRECISE NUMBER

Adults, of course, have had a lifetime of language experience, and so their intuitions do not yet inform our understanding of what would compel a child to decide what meaning the speaker of the sentences in (1) or (2) had in mind. Would children entertain the same restricted set of properties as possible meanings for *gleeb* in (1) as adults do, or would they allow for more possibilities? Understanding what children do when given novel words like *gleeb*, and under what linguistic and extralinguistic circumstances, would aid researchers in finding the unique biases and expectations that allow children to acquire words for numerical (*quantity*) and non-numerical (*quality*) properties.

Investigating the acquisition of any linguistic phenomenon demands consideration of the roles of the basic components of language acquisition (Lidz & Gagliardi 2015). This means looking at what is in the *input*, the raw data available to the learner, and distinguishing this data from the *intake*, or, the portion of the input that the learner is able to make use of at any given stage of language development. In addition to these factors, we need to consider how the learner's hypothesis space is shaped, both by potentially innate language specific hypotheses, as well as those governed by the grammar and lexicon that the child has acquired so far. Finally, we need to understand what kinds of inferences a learner *could* make to determine which of the hypotheses in this space are supported by the data in the intake, and so should be generalized to the grammar. In this paper, we look at how these components interact to determine children's acquisition of novel superlative words like *gleebest*.

§2 explores syntactic bootstrapping as a possible mechanism for the acquisition of superlative words. As a theory of word learning, syntactic bootstrapping is predicated on three things: (i) that there exist principled distributional cues to learners; (ii) that those cues are usable in learning; and (iii) that children use those cues at the initial stages of learning. The distributional cues that we consider are grammatical categories. As we explain below, when a novel superlative is used in a

determiner position (e.g., *Gleebest of the cows are by the barn*; cf. *Most of the cows are by the barn*), knowledge of the syntax-semantics mapping for determiners constrains the possible space of meanings to those with quantity-based interpretations, and no further inference is required. When the superlative is found in an adjectival position (e.g., *The gleebest cows are by the barn*; cf. *The most/reddest cows are by the barn*), however, the set of possible meanings is less constrained: both quantity- and quality-based meanings are available. We show that 4 year old children can use these cues, establishing (i) and (ii), and setting the stage for investigating (iii) (cf. the strategy of Piccin & Waxman 2007, and the Human Simulation work beginning with Gillette et al 1999).

Use of the superlative allows for a direct comparison between the hypothesis that partitivity is a strong cue to quantity-based meanings (Syrett, Musolino & Gelman 2012) and the hypothesis that, rather, it is parsing a novel word as occurring in a Determiner (D) position. We present novel experimental results supporting the latter hypothesis. To preview: when *gleebest* appears in a D position, children choose a quantity-based interpretation, but they choose a quality-based interpretation when the superlative word occurs in an Adjective (A) position, regardless of whether it is paired with a partitive noun phrase. We use these data to argue that four year old children know and deploy knowledge of the syntax-semantics mapping for determiners, as adults do. These findings contribute to the literature on preschoolers' rich early knowledge of the syntax-semantics mapping for determiners, as revealed through novel word learning paradigms (Hunter, Lidz, Wellwood, & Conroy 2011, Hunter & Lidz 2013).

This investigation leaves open the question of what explains the strong bias children show towards quality-based meanings when the novel word occurs in an A position. Since words appearing here may have either quantity- *or* quality-based meanings in the adult grammar, we may have expected ambivalent responses in this case. In §3, we consider four computational models in an effort to better understand this bias. Using Bayesian inference we are able to probe the relative contributions of a learner's prior expectations about potential word meanings, and their ability to reliably encode different kinds of information from the word-learning context. These models combine the expectations pulled from the learner's lexicon (i.e., the distribution of quantity- and

quality-based meanings for words that can appear in A positions) and the perceptual differences affecting how the qualities and quantities manipulated in the experiment are encoded. The results of our simulations reveal that the learner’s prior expectations about word meanings are the driving factor in the generalizations children make when they see a novel word in A positions. By including this investigation we are able to give a fuller picture of novel word meaning inference: we know both how syntax can govern children’s generalizations about novel word meanings, and which factors come into play when syntax does not tightly constrain the space of hypothesized meanings.

2 Syntactic bootstrapping

The best case for a syntactic bootstrapping explanation for how children acquire novel property-denoting words is in the case of words with quantity-based meanings, since there are syntactic environments in which these expressions can occur but words with quality-based meanings cannot. One such environment is the syntactic position of a determiner: crosslinguistically, words appearing here can only express quantities of things, never qualities of things. If the learner is able to identify what counts as a D in their language, then the meaning of a novel word presented in this position will be appropriately restricted. In this section, we discuss this hypothesis, and then present the results of our word learning experiment.

Recently, Syrett, Musolino and Gelman (2012) tested the hypothesis that the partitive frame (i.e., *of the cows*) is a strong cue to quantity-based meanings. If this hypothesis is correct, then embedding a novel word in this frame should lead children to pick a quantity-based interpretation in cases when both this and an alternative, quality-based interpretation are available. In Syrett et al.’s word learning task, they restricted the potential referents for the novel word *pim* to the quantity TWO and the quality RED.² They found that the partitive predicted quantity-based judgments only in restricted cases,³ casting doubt on the robustness of a syntactic bootstrapping account based on

²We follow custom in using italics for linguistic expressions and small caps as shorthand for their meanings.

³Only when it was used at test; when the partitive was used during training but not at test, children were at chance

the partitive as a strong cue. Presenting the results of a novel corpus study, they point out that a great variety of non-quantity-referring expressions occur in this frame (3), suggesting that perhaps we should not expect the partitive to be a strong cue to quantity-based meanings.

- (3) a. **Amount:** *all, two, seven, most, some*
b. **Segment:** *back, front, edge, side, top*
c. **Measure:** *mile, hour, pound, bucket*

So far, then, the puzzle raised by Wynn's (1992) original finding remains: how is it that children at 2;6 could know that number words express quantity, despite not knowing which quantities?

We think an alternative explanation of Syrett et al's results is available. First, we agree that the partitive is not likely to be a strong cue: it is too language-specific, and consequently to learn *that* it is a cue would first require knowing what it is a cue for. A stronger cue is whether something occurs to the left of *X* in *X of the NP*. Of the classes of counterexamples provided by Syrett et al given in (3), only the "Amount" terms can appear without a determiner (e.g., *a* or *the*) on the left:

- (4) a. Two/most of the cows have been milked already.
b. * Back/side of the fridge is heating up dangerously.
c. * Mile/hour of the race was uphill.

Such data illustrate an important linguistic generalization: words expressing quantities can appear in the syntactic position of determiners.⁴ Unlike the broad class of expressions that can appear in the partitive frame, determiners have a stable syntax-semantics mapping cross-linguistically: their interpretations only reference quantities, never qualities, of individuals (i.e., they have the property of *permutation invariance*; Barwise & Cooper 1981, van Benthem 1989, Gajewski 2002). Observing this pattern leads to the following hypothesis: if a child categorizes a novel word as D, she will understand that word to have a quantity- rather than quality-based meaning.

at picking the quantity interpretation.

⁴As a simple rule to determine which word in a string is D, take *X* in *X of the cows* to be D unless *the* precedes *X*. Since *the* cannot appear without an element to its right before *of* (cp. **the of the cows*), it instantiates D whenever it is present. In *the most cows*, *the* instantiates D, but *most* instantiates D in *most of the cows*.

Given these considerations, we are left with a puzzle: Syrett et al in fact presented the novel word *pim* in a determiner position in some conditions, with their target question taking the form *Who has pim of the trains?* If determiner status were a strong cue to quantity-based meanings, why would they get mixed results in those conditions? We think this is because the quantity-based meaning that Syrett et al licensed was the precise numerical concept TWO. The mean (and median) age of children in their study was 3;10, which is slightly before the age children generally come to understand the Cardinality Principle (see Carey 2010 for an overview). Given that it is still unknown why children have such difficulty in acquiring words for precise number, a reasonable hypothesis is that isolating this concept in Syrett et al's task was too difficult for children of this age. Indeed, Halberda, Taing, and Lidz (2008) showed that many children comprehend expressions like *most* prior to precise cardinality words. Thus, it might be necessary to understand how children decide that novel words express quantity notions at all before we can understand how they learn the meanings of words for exact number (see also Barner, Chow & Yang 2009 for discussion).

In our experiment, we test whether partitivity versus syntactic category is a strong cue to quantity-based meanings. Like Syrett et al, we consider an appropriate test to be one that makes both quantity- and quality-based properties salient, and measures children's interpretive preferences as a function of the novel word's syntactic context. However, we tested children with a novel superlative word *gleebest*. As we will see, superlatives uniquely allow for the opportunity to directly compare the hypothesis that syntactic category as opposed to partitivity is a strong cue to quantity-based meanings. More importantly, however, they allow us to avoid the problem just discussed: superlatives express relative as opposed to absolute quantity, and so may allow children to isolate quantity as a relevant dimension without having to further determine a specific quantity.

Combining a word like *heavy* with the morpheme *-est* allows the formation of expressions like *the heaviest animals*, with a meaning like THE ANIMALS THAT ARE HEAVIER THAN ANY OTHERS. Similarly, combining *many* with *-est*⁵ gives *the most animals*, with a meaning like THE

⁵More specifically, *most* is the superlative of *many* or *much*, see Bresnan 1973, Hackl 2009; see Bobaljik 2012 for an alternative analysis in which it is the superlative of *more*.

ANIMALS THAT ARE MORE NUMEROUS THAN ANY OTHERS. Importantly for our purposes, both of these types of superlatives can surface in the position of an adjective (5a) (where *the* instantiates the syntactic category D), but only the quantity-based superlative *most* can appear bare on its left, itself instantiating D (contrast (5b) with (5c)):

- (5) a. The heaviest/most animals are happy.
- b. Most of the animals are happy.
- c. * Heaviest of the animals are happy.

Notice that such a contrast as is observed in (5b-5c) cannot be conceptual in nature, and thus must relate to the grammatical status of determiners. While (5b) can be paraphrased MORE THAN HALF OF THE ANIMALS BY NUMBER ARE HAPPY, by straightforward analogy a possible meaning of (5c) should be MORE THAN HALF OF THE ANIMALS BY WEIGHT ARE HAPPY. To see what the latter would mean, consider a situation in which the only animals are a cow *C*, a lamb *L*, and a rabbit *R*. (5b) here would be judged true if any two of these animals are happy. But the proposed meaning of (5c) would require more information than just how many animals there are: if *C* weighs 700kg, *L* weighs 35kg, and *R* weighs 8kg, (5c) would be judged true as long as (at least) *C* is happy. *C*'s weight is so great that (5c) can be true if he alone is happy, or if he and *L* are happy, and so on, but is false if *C* is not happy. Individuals and their particular properties matter for quality-based superlatives (i.e., they are **not** permutation invariant), where only set cardinality matters for *most*. While clearly no conceptual necessity rules out such a meaning for a quality-based adjective like *heavy*, such form-meaning pairings are not observed across the world's languages.⁶

Finally, a critical property of superlatives is that, regardless of whether they have a quantity- or quality-based meaning, they can appear in adjectival position with the partitive frame, (6a-6b). The root expressions of *most* and *spottiest*, *many* and *spotty*, do not have this property, (6c-6d).

- (6) a. The spottiest of the cows were by the barn.

⁶This is especially surprising in light of recent proposals in the formal semantics literature that, semantically, *most* and *heaviest* are indistinguishable (Hackl 2009). Yet, it is difficult to see how an appeal to a property like numerosity would support the formulation of a *syntactic* constraint that could make sense of the facts in (5a)-(5c).

- b. The most of the cows were by the barn.
- c. * The spotty of the cows were by the barn.
- d. * The many of the cows were by the barn.

In the next section, we put our hypothesis that syntactic category cues quantity-based meanings to the test in a novel word learning task, contrasting this hypothesis with that suggesting partitivity is a strong cue.

2.1 Experiment

In the previous section, we hypothesized that representing a novel word as an instance of the category D was a strong cue to the learner that the word should be assigned a quantity-based meaning. An alternative was presented that suggested presence of the partitive frame alone was a strong cue. We test these ideas by examining children’s preferences when embedding the novel superlative word *gleebest* in a variety of syntactic contexts, using a variant of the Picky Puppet task (Waxman & Gelman 1986; see Hunter, Lidz, Wellwood, & Conroy 2010, Hunter & Lidz 2013 for extensions to novel determiners).

Method. In this task, the experimenter first explains to the child that the name of the game is to sort cards according to whether a puppet likes them or not. The child is told that the puppet is picky, but is usually friendly enough to share the reasons for his preferences. The experimenter explains the puppet’s criterion for a given set of cards by saying: “For these cards, the puppet said he likes the cards where **target sentence**, but he doesn’t like the ones where it’s not true that **target sentence**”. A subset of the cards are sorted for the child by way of demonstration, and then the remaining cards for a given set are handed to the child one at the time, and the child puts them either in the “likes” pile (marked with a green checkmark) or the “dislikes” pile (marked with a red X) depending on how they interpret the target sentence.

Three short warm-up games of this form are played first, each comprised of a set of 6 cards (3 true, 3 false) and a distinct target sentence. In the first warm-up game, the target sentence is

“The puppet likes the cards where everything’s red.” This game ensures that children can respond to the contents of the cards holistically. In the second warm-up game, the target sentence is “The puppet likes the cards where there are more hats than t-shirts”. This game ensures that children can compare subsets of items on a card to one another. In the third and final warm-up game, the target sentence is “The puppet likes the cards where everything’s *blick*”. Here, the child is additionally told that in this case the puppet was being silly, and wouldn’t tell the experimenter what *blick* meant (which was PURPLE) but maybe the child could help the experimenter figure it out. This game ensures that children would not balk when presented with a novel word. Our participants received no instructive feedback during these warm-up games, yet had no difficulty sorting the cards correctly.

[FIGURE 1 ABOUT HERE.]

The experiment itself proceeded in two phases. In the Training phase, the experimenter first introduces the child to what information is relevant from the new set of cards, saying: “These ones are all kind of similar. There are some cows (pointing first to one group and then another), a field (pointing to the field behind one group of cows), and a barn (pointing to the barn behind the other group of cows). For these cards, the puppet said he likes the ones where **determiner phrase (DP)** are by the barn”, where the DP was manipulated between subjects. Each DP contained the novel word *gleebest* in either adjectival (ADJECTIVAL), confounded (CONFOUNDED), or determiner (DETERMINER) positions, as determined by different combinations of the presence/absence of *the* and the partitive frame (Table 1). The experimenter then explains that the puppet was being silly again and wouldn’t tell her what *gleebest* means, but was hoping the child could help her figure it out. The child is then shown 6 training cards one at a time (Figure 1), the ones the puppet had “already told” the experimenter it liked or didn’t like, appropriately sorted into the “likes” and “dislikes” piles.

[TABLE 1 ABOUT HERE.]

In the Test phase, children are handed 12 cards one at a time, and asked whether they think the puppet likes that card or not. While the training cards are perfectly ambiguous (the group by the barn is both the most numerous and the most spotty), the test cards are perfectly *unambiguous*. For our test cards, the ratio of the numerosities of the cows was inversely proportional to the ratio of the spots of the cows, see Table 2. The same cards (in counterbalanced order) were presented to each participant. The experimenter handed each test card to the child with prompts like “Do you think he likes this one?”, “What about this card, do you think he likes it?”, and the child was to place the card either in the “likes” or “dislikes” pile. Training cards remained visible above the corresponding piles throughout the experiment. At the end of the experiment, the child was probed as to what they thought *gleebest* meant, and responses were recorded.

[TABLE 2 ABOUT HERE.]

[FIGURE 2 ABOUT HERE.]

We hypothesized that categorizing a novel word as a determiner would restrict a child’s interpretation of the meaning of *gleebest* to a quantity-based meaning. An alternative hypothesis was that the presence of the partitive frame was a strong cue to such interpretations. The relevant hypotheses are schematized in Table 3 according to whether they predict a greater-than-chance quantity-based response (indicated by +).

[TABLE 3 ABOUT HERE.]

36 children participated (range 4;0-5;2, mean 4;7). Each child was given a small gift for participating. Four additional children were tested and subsequently excluded—2 due to experimenter error, 1 due to presenting with a strong “yes” bias (i.e., the participant indicated the puppet “liked” 11/12 of the test cards), and 1 due to a strong “no”-bias (i.e., they said the puppet “didn’t like” 12/12 of the test cards). We measured the percentage of cards sorted consistent with a quantity-based interpretation.

Results. Across our three conditions, sign tests showed that responses were significantly different from chance (ADJECTIVAL $Z = 5.48, p < .001$; CONFOUNDED $Z = 2.37, p < .05$; DETERMINER $Z = 5.93, p < .001$). These differences were in different directions, however. Children sorted cards consistent with a quantity-based interpretation in DETERMINER 72% of the time, compared to 29% in ADJECTIVAL and 40% in CONFOUNDED. In addition, DETERMINER was significantly different from both ADJECTIVAL, $t(22) = 3.03, p < .01$, and CONFOUNDED, $t(22) = 2.20, p < .05$. These results are presented graphically in Figure 3. It is noteworthy that 8 out of 12 of the children in DETERMINER sorted at least 9 out of 12 test cards consistent with a quantity-based interpretation, while only 2 out of 12 children did so in ADJECTIVAL and only 3 out of 12 in CONFOUNDED. Children’s responses to the question “What do you think *gleebest* means?” following the experiment confirmed our interpretation of these results, using phrases like “more spots” or “more cows” (as appropriate) when paraphrasing their understanding of the word.

[FIGURE 3 ABOUT HERE.]

As there were no differences between our conditions except for the syntactic context in which *gleebest* occurred, these results support the hypothesis that syntax acts as a strong cue guiding children into quantity-based meanings. However, our manipulations suggest that syntactic category is a stronger cue than partitivity is. There was no effect of partitivity, i.e. CONFOUNDED was no different than ADJECTIVAL, $t(22) = .72, p = .48$. In both conditions, children displayed *lower* than chance sorting of cards consistent with that interpretation. Of the three hypotheses sketched, only syntactic category as a strong cue captures the results we found.

2.2 Discussion of experimental results

Our results show that 4 year-old children use the syntactic position of a novel superlative to assign either a quantity or quality-based interpretation: children sorted cards consistent with a quantity-based interpretation for *gleebest* only when it occurred in the syntactic position of a determiner. In addition, the results show that the presence of the partitive *of* is not a strong cue to quantity-

based meanings: children sorted cards consistent with a quality-based meaning for *gleebest* when it occurred between a determiner and the partitive *of*. These results are important for a number of reasons.

First, as discussed in the introduction, choosing “quantity” as the relevant property from a set of available properties is potentially challenging for children. Our use of a novel superlative word *gleebest* may have made this task easier than would the use of a non-superlative word, since the quantity-based meaning it suggests is not absolute (such as the quantity TWO), but rather comparative (MORE THAN another quantity): all the child needs to do is figure out that quantity is the intended dimension, and not the further step of exactly *which* quantity. This gets us part of the way to understanding Wynn’s (1992) finding, that children interpret number words as expressing cardinality before they understand that they express particular quantities. Children use their knowledge of the syntax-semantics mapping of determiners to restrict the hypothesis space of possible meanings.

An additional but related question that this work raises is the strength of the bias towards quality-based meanings in ADJECTIVAL. (Since ADJECTIVAL and CONFOUNDED did not differ from one another, we focus now only on the bias observed in ADJECTIVAL.) Given that children had no problem deciding that *gleebest* had a quantity-based meaning when it was presented in determiner position, we cannot assume some inability to reason about number when presented with that expression in a non-determiner position. A reasonable speculation is that the bias observed in ADJECTIVAL is due to the child’s distribution of known superlative meanings: if many more words in this category refer to properties of objects than to properties of sets, the prior distribution of meanings of words in this category could bias the child towards the former kind of meaning, absent syntactic cues to the contrary. The next section introduces computational models designed to investigate the source of this bias.

3 Beyond syntactic bootstrapping

We saw in the preceding section that while syntactic bootstrapping can account for the preference for quantity-based meanings for a novel superlative when it is presented with the syntax of a determiner (which only allows for such meanings), it is not enough to sufficiently constrain the set of available meanings when such a word appears in an adjectival position. Despite the fact that syntax leaves the choice between quantity- and quality-based meanings open in this case, it was observed that children nonetheless showed a preference for quality-based meanings. That is, when grammatical category constrains the meaning (as in the DETERMINER condition), we see children picking the meaning allowed by the syntax; when grammatical category does not constrain the meaning (as in the ADJECTIVAL condition), children still show a bias in their preferences. In this section, we use computational modeling to examine the nature and source of this bias.

The bias towards quality-based meanings when a novel superlative appears in adjectival position could stem from two sources in the learner: (a) lexical bias, based on the distribution of known superlative (or, gradable adjective) meanings in the child’s lexicon, or (b) salience, the relative difficulty of perceiving and encoding the differences in the numerosity of groups of cows versus their relative levels of spottiness. As word learning in ambiguous contexts has been successfully modeled using Bayesian inference (Xu & Tenenbaum 2007, Gagliardi et al 2012), we adapt these models to explore how (a) and (b) could interact when children make generalizations about the meanings of novel superlatives appearing in adjectival position.

Xu and Tenenbaum (2007) showed that Bayesian inference could be used to accurately model children’s performance on inferring novel object labels. In particular, they showed that children can use expectations about the size of the set of objects being labeled to determine the most likely meaning of the novel word. Gagliardi et al (2012) showed that by adding children’s knowledge of grammatical category, and their expectations about what kinds of concepts a given grammatical category tends to express, children’s performance on both novel noun and adjective learning can be predicted. That is, children have different expectations about likely meanings for novel nouns and novel adjectives, and these expectations can be modeled by looking at the distribution of concept

types across these categories in the developing lexicon.

In what follows, we use the same kind of inference model, taking into account the fact that children have access to the syntactic knowledge that the novel word presented to them in our experiment occurred in adjectival position. This allows us to look at the specific biases that might be tied to knowledge of the likely meanings of novel superlative adjectives.

3.1 Bayesian inference

Bayesian inference calculates the posterior probability of a hypothesized meaning, h , given some observed data, d . Here, the hypotheses are whether a novel word encodes a quality- or quantity-based meaning. The posterior probability, $\mathbb{P}(h|d)$, is proportional to the product of the prior probability of each hypothesis, $\mathbb{P}(h)$, and the likelihood of each hypothesis given the data, $\mathbb{P}(d|h)$ (Equation 1).

$$\mathbb{P}(h_i|d) = \frac{\mathbb{P}(d|h_i) \cdot \mathbb{P}(h_i)}{\sum_{h_j \in \{\text{all hypotheses}\}} \mathbb{P}(d|h_j) \cdot \mathbb{P}(h_j)} \quad (1)$$

By using Bayesian inference we can directly probe the role of the learner's prior beliefs about each hypothesis ($\mathbb{P}(h)$) in their generalizations about a novel superlative's meaning. Additionally, by using a *mixture model* that combines the probabilities of making different inferences about the same data, we can investigate the role played by their ability to encode the relevant features of the pictures expressed by the novel superlative.

3.2 Four models of inference

As stated above, the aim of this investigation is to uncover the role played by (a) lexical bias, based on the distribution of superlative adjectives in the learner's existing lexicon and (b) salience, the relative difficulty of perceiving and encoding differences in numerosity versus spottiness of the groups of cows. To this end, we built four models, each representing a different combination of these factors into the inferences used to generalize novel word meanings. Two simple models

encode only lexical bias (Lexical Bias - Model 1) or conceptual bias (Conceptual Bias - Model 2) directly into the prior. One slightly more complex model has these factors both influence the prior (Lexical and Conceptual Bias - Model 3). A different, but perhaps more realistic model (Perceptual Bias - Model 4), assumes that while lexical bias influences the prior, salience acts as a confusability parameter. In each model, the likelihood, $\mathbb{P}(d|h)$, is assumed to be equal for each hypothesis, as our training stimuli were designed to make both numerosity and spottiness equally good fits for description with *gleebest*.

Model 1, the Lexical Bias model, looks at what kinds of generalizations the learner would make if only lexical bias influenced their inferences. We first calculated the prior probability of each kind of novel adjective (quantity-based and quality-based) based on a hypothetical child's lexicon. This "lexical prior", $\mathbb{P}(h^{lexicon})$, was approximated via a count of gradable adjective types from parental speech in four CHILDES corpora (Adam, Eve, Sarah, Nina; MacWinney 2000). Syrett (2007) isolated 45 quality-based gradable adjective type in this corpus. We searched the same corpus and found 5 quantity-based gradable adjectives (*more, many, much, few, little*). We focused on gradable adjectives as these are the words that may combine with *-est* to form a superlative word; counting only unique occurrences of superlatives would likely have dramatically underestimated children's lexical knowledge. The resulting approximation is nonetheless quite conservative, as the 4-year-old participants in our experiment likely had larger lexicons. However, given the very small number of adjectives with quantity-based meanings even in the adult lexicon, any increase in the size of the children's lexicons would only be in the number of quality-based adjectives. This means that, if our approximation of the lexicon is skewed, it is skewed in the direction of making quantity-based meanings more probable than they would otherwise be, a factor worth remembering when analyzing the predictions of this model.

Following Gagliardi et al (2012), we use the counts of adjective types from the lexicon to calculate the the lexical prior ($\mathbb{P}(h^{lexicon})$) as a beta binomial distribution equivalent to Equation 2, where α is equal to the number of gradable adjectives of either quantity or quality-based adjectives in the lexicon, and β is equal to the total number of gradable adjectives.

$$\mathbb{P}(h_{lexicon}) = \frac{\alpha + 1}{\alpha + \beta + 2} \quad (2)$$

Thus the resulting Lexical Bias model is that shown in Equation 3.

$$\mathbb{P}(h_i|d) = \frac{\mathbb{P}(d|h_i) \cdot \mathbb{P}(h_i^{lexicon})}{\sum_{h_j \in \{\text{all hypotheses}\}} \mathbb{P}(d|h_j) \cdot \mathbb{P}(h_j^{lexicon})} \quad (3)$$

Model 2, the Conceptual Bias model, ignores lexical statistics, inferring meanings based only on how salient the differences in numerosity and spottiness are, where $\mathbb{P}(h_{salience})$ is the probability of either dimension being singled out, and hence perhaps lexicalized. To derive this, we approximated a “salience prior”, $\mathbb{P}(h_{salience})$, based on a similarity rating study we conducted on 50 undergraduates who received course credit or \$10 for participating. Subjects were presented with pairs of pictures differing along only one dimension (either pairs of single cows with 1-10 spots [quality], or pairs of groups numbering 1-10 [quantity]) on a computer screen, and asked to rate the images based on their similarity (1=not at all similar, 9=very similar) (Figure 4). Each subject saw each pairing of each cow and spot comparison twice. The order of presentation and screen position of each picture in a pair (right or left) was randomized using MATLAB.

[FIGURE 4 ABOUT HERE.]

Based on hierarchical clustering of the resultant similarity judgments, we measured cluster distinctiveness within the two dimensions (cf. Xu & Tenenbaum 2007, Gagliardi et al 2012). As can be seen in Figure 5, the differences in mean similarity judgments for distinct ratios are larger for quantity-based differences than quality-based ones. We take this to mean that differences in the quantity of cows across cards are more salient than differences in the spottiness of cows. To quantify these differences, we took the measures of mean cluster distinctiveness for each dimension to be indicative of how salient distinctions among concepts in either dimension are, and thus how accessible they might be to the process of lexicalization. This is an admittedly indirect measure of something that may not be quantifiable, but we believe it can serve as an approximation of the

difference in salience between these two dimensions. The mean cluster distinctiveness for quantity and quality based comparisons which we used as $\mathbb{P}(h^{salience})$ are shown in Table 4, and the resulting Conceptual Bias model is shown in Equation 4.

[FIGURE 5 ABOUT HERE.]

[TABLE 4 ABOUT HERE.]

$$\mathbb{P}(h_i|d) = \frac{\mathbb{P}(d|h_i) \cdot \mathbb{P}(h_i^{salience})}{\sum_{h_j \in \{\text{all hypotheses}\}} \mathbb{P}(d|h_j) \cdot \mathbb{P}(h_j^{salience})} \quad (4)$$

Model 3, the combined Lexical and Conceptual Bias model, used a “complex prior” for each hypothesis, which is the joint probability of the lexical and salience priors. This model allowed us to look at what generalizations would be predicted if both lexical bias and the salience of a concept were tied to the prior probability of each hypothesis.

Model 4, the Perceptual Bias model, took a different approach, combining the lexical prior with the intuition that salience impacts how the likelihood, $\mathbb{P}(d|h)$, could be encoded with differing reliability for each hypothesis (cf. Gagliardi & Lidz, 2014). As stated above, the likelihood of the data is the same for every model, but this model manipulates whether this likelihood is even computable for a given observation. If the learner can't encode the relevant data, the computed likelihood will be different from when they can. In other words, this model takes into account whether or not the learner can reliably encode the relevant data needed to support each hypothesis. For example, the learner could have trouble perceiving, and hence encoding, the numerosity of the groups of cows in a picture. This would mean that data for the quantity-based hypothesis of *gleebest* would be unclear, but data for the quality-based meaning would remain apparent. Alternatively, the learner could be proficient at encoding numerosity but not at reliably encoding spottiness. Finally, it is possible that the learner would be able to reliably encode both spottiness and numerosity, or not be able to reliably encode either. The probabilities of misencoding numerosity or spottiness are

incorporated as two free parameters into the Perceptual Bias model (α and β , respectively). The probability of misencoding both numerosity and spottiness at once (if we assume these probabilities to be independent, is the joint probability of the two (γ).

Applying these two free parameters to our combined model (Model 3) yielded four terms altogether (A, B, C and D), combined in a mixture model (the sum of all four terms). Each of these is the posterior probability of the two hypotheses, given the encoding of the data, where each term was multiplied by the probability of encoding each data type (Table 5). It is important to remember that, in Model 4, α and β are free parameters correlated with the relative salience of distinctions on the quantity and quality dimensions. In the simulations that follow, we set these to be consistent with the relative confusability of spottiness versus numerosity, as measured in our similarity judgment task. That is, we hypothesize that spottiness is more confusable than numerosity, and set the values to be in line with the differences found for cluster distinctiveness (Table 6). However, we set these values manually as we have not yet determined the appropriate transform from similarity (which we measured) to confusability (which we need here).

[TABLE 5 ABOUT HERE.]

[TABLE 6 ABOUT HERE.]

All four of our models, along with the role played by the lexical bias and salience in each of them, are summarized in Table 7.

[TABLE 7 ABOUT HERE.]

3.3 Simulation results and discussion

The results of the simulations are seen in Figure 6, which shows the posterior probability of each hypothesis predicted by the model. Only the models that incorporate the biases from the lexicon (i.e., Models 1, 3 and 4) reflect the general pattern exhibited by the children in the experiment

(i.e., the fourth column in Figure 6). Model 4, the Perceptual Bias model, provides the closest fit, suggesting that children’s generalizations about novel word meanings could be a function **both** of the biases they bring to the word learning task and *their ability to reliably encode information* in the world. However, it is important to remember that the two free parameters in this model allow for an arbitrarily close match to the children’s data. While we set these parameter values to be consistent with the apparent confusability of spottiness and numerosity, they were not direct transforms of the measured similarity between differences on these two dimensions.

[FIGURE 6 ABOUT HERE.]

In general, it is not clear that we need a perfect fit of the model to the data. That is, in the Perceptual Salience model we could be overfitting experimental noise instead of fitting children’s inferences. To understand this point, recall the DETERMINER condition in the experiment, where children were presented with *gleebest* in determiner position. This syntactic context only permitted a quantity-based interpretation of the novel superlative, and while children clearly preferred this interpretation, they sometimes chose the quality-based interpretation; this is most likely due to some kind of experimental noise. In the ADJECTIVAL condition, when *gleebest* was presented in adjectival position, children could be as strongly biased as the Lexical Bias model alone suggests, but again experimental noise might be the reason that their behavior doesn’t perfectly match the predictions made by the model. Of course, ‘experimental noise’ isn’t an explanation in and of itself; the source of this noise is worth considering, and it could be caused in part by the salience of the relevant properties. That is, it could be that this is exactly what we are modeling in the Perceptual Bias model. This concern aside, what remains clear from our simulations is that Lexical Bias accounts for the major trend seen in children’s generalizations for novel superlative adjectives.

Before concluding our discussion of the source of the quality-based preferences in the ADJECTIVAL condition, it is worth discussing and ruling out two further possible sources of this bias.

First, one might be inclined to argue that given a constrained set of hypotheses, the likelihood of finding the same construction used to convey a given meaning over and over would drive to learner

to infer that the meaning associated with fewer possible structures is the more likely one (cf. Xu & Tenenbaum 2007). In other words, if each possible meaning of *gleebest* (limiting ourselves to the options SPOTTEST and MOST NUMEROUS) has a constrained set of structures that could be used to convey the meaning (say, the structures employed in the experiment), then the meaning associated with the smaller number of possible structures becomes more likely. Recall that all three structures used were compatible with the quantity-based interpretation, but only the structures where *gleebest* did not occur in determiner position were compatible with the quality-based meaning. This would mean that the likelihood of the latter structures given the quality-based meaning would be one out of two (0.5; there were only two structures to choose from so there was a probability of 0.5 of picking either one), and their likelihood given the quantity-based meaning would be 1 out of 3 (0.33; there were three structures to choose from for this meaning, so a probability of 0.33 of picking any one). In the absence of any other biases, since 0.5 is greater than 0.33 this means that the learner should prefer the quality-based meaning in the ADJECTIVAL condition. This kind of difference in likelihood is compounded by the fact that the learner heard the puppet make the same choice six times during training (which diminishes the likelihood exponentially, magnifying the differences), and is thus more than enough to strongly bias the learner towards the quality-based meaning in the ADJECTIVAL condition.

However, this powerful inference process depends directly on the learner entertaining only a highly constrained set of possible structures as ways of conveying the meaning contained in the experimental utterances. In reality, it is unlikely that the learner would only consider just the three possible structures we chose for the experiment as the possibilities for expressing the intended meaning. We know of no principled way to constrain the set of possible structures that a learner might entertain for a given meaning, given that the syntax ultimately allows for boundlessly many possibilities. And without a constrained set, it is difficult to see how the sizes of the sets of possible structures for each meaning could be compared, and thus it is not possible to probe the inferences that such comparisons might drive.

Second, another way of thinking about the quality-based meaning preference in the ADJEC-

TIVAL condition considers that the ratios of spots on the cows on the experimental training cards were always smaller (hence plausibly, easier to distinguish) than the ratios between the numbers of cows. For example, one card had a 2:1 (0.5) ratio of cows in the field and by the barn, while it had a 13:1 (0.07) ratio of spots. It seems possible that such differences (see Table 2 for all ratios used in training) highlighted the differences in spottiness. Indeed, considering the mean similarity ratings obtained in our similarity judgment experiment, we can clearly see that mean similarity score increases as ratios become harder to distinguish (Figure 7). The data shown in Figure 7 could lead us to believe that the contrasts on the dimension with the easier ratios in training (quality) were easier to perceive than the dimension with the more difficult ratios (quantity), and that this difference alone caused the quality meaning to be more salient, and thus preferred in the syntactically uninformative conditions.

[FIGURE 7 ABOUT HERE.]

This possibility can be dispelled by looking more closely at the results of the similarity judgment experiment. The disparity in ratios between number of spots and number of cows would only highlight the differences in spottiness if differences in spottiness and numerosity align with one another as the ratios change. However, when we distinguish between judgments made on ratios of spottiness and those on ratios of numerosity, the data suggest that differences between numbers of spots and numbers of cows do not consistently align with one another. That is, given the same ratio between different numbers of cows on two cards, or different numbers of spots, participants judged the cards with differing numbers of cows to be ‘more different’, than those with differing numbers of spots. This is illustrated in Figure 8, which plots the mean similarity rating for each judgment as a function of the ratio, but keeps the condition (quantity versus quality) distinct. We can see that while both types of comparison are judged as ‘more similar’ as the ratio gets larger (harder to discern), the similarity scores for numbers of cows are reliably different than those for numbers of spots. Moreover, if we look at the points on this graph corresponding to the training ratios (colored numbers), while the ratios for the number of spots on the training cards are reliably

lower (red numbers), the projected similarity judgements between these ratios and those of the numbers of cows on the corresponding training cards (blue numbers) are not. For the ratios used in Trial 1, the similarity judgments are roughly equal; for those used in Trial 2, the difference in spottiness appears to be more salient ('less similar'), and for Trial 3 the difference in numerosity appears to be more salient ('more similar').

[FIGURE 8 ABOUT HERE.]

Of course, one could argue that since the similarity judgment experiment was performed on adults, we cannot draw conclusions from it about children. It could be that, for children, the ratios used in training really did make the difference in spots more salient than the difference in numerosity. While it is possible that adults' judgments about similarity of numbers of spots versus numbers of cows differ from those of children, we have no reason to suspect that this would be the case. Were a plausible hypothesis put forward as to why children might differ from adults in this way, we would then be forced to reevaluate the claims made here.

4 Conclusion

How children acquire words as quickly and as apparently effortlessly as they do is a major question in language acquisition research, and one that requires considering a number of different factors, both linguistic and extralinguistic. The first step for the researcher is to appreciate the difficulty of the task in the case of some types of meanings. In this paper, we focused on words for qualities (like *spotty*) and for quantities (like *most*). After this step, we require careful consideration of what information in the environment is available to children's inferences (the *data*) but also of what their current cognitive and linguistic abilities allow them to actual filter from this data (their *intake*). Compounding the issue is the fact that children's grammars and lexicons are changing over the course of acquisition, and this information can be fed back into the inferences they engage later on.

In our consideration of the acquisition of a novel superlative word, we were able to see the contribution of each of these components. When children were presented with *gleebest* in our word-learning experiment, their grammatical knowledge allowed them to accurately encode what structural position this novel superlative appeared in. From this encoded intake, they could determine that the novel word was either a determiner or an adjective. When the word appeared in the syntactic position of a determiner, they knew that only one kind of meaning was available—a quantity-based meaning. When *gleebest* appeared in the position of an adjective, either a quantity- or quality-based meaning was available, yet we found that children preferred a quality-based interpretation. Computational modeling revealed that while this preference is driven by statistics drawn from the lexicon, the learner’s ability to reliably encode relevant property distinctions (i.e., the relative differences in numerosity or spottiness across groups of cows) could also influence this inference, or its availability.

Understanding what these models represent, coupled with the experimental results we report, highlights the importance of the linguistic knowledge that children bring to the word learning task (lexical *and* syntactic), as well as the extralinguistic capacities and limitations inherent to the learner. Finally, these results emphasize the contributions that computational modeling, combined with careful empirical work with both adults and children, can make to our understanding of both language acquisition and linguistic representations.

5 Acknowledgments

This work was made possible by generous support from a Social Sciences and Humanities Research Council of Canada doctoral award (#752-2010-0499) to A. Wellwood, an NSF GRF to A. Gagliardi, and NSF IGERT 0801465. The authors would especially like to thank Justin Halberda, Darko Odic, Paul Pietroski, and Naomi Feldman, as well as Tim Hunter, Research Assistants Leah Whitehill and Jessica Lee, the University of Maryland’s infant and preschool labs, the Center for Young Children, and the audiences at the Linguistic Society of America’s 2012 annual meeting,

the Cognitive Science Society's 2012 annual meeting, and the North Eastern Linguistics Society 2012 annual meeting.

Figure 1: Ambiguous training cards, sorted according to whether the puppet “likes” them or not.

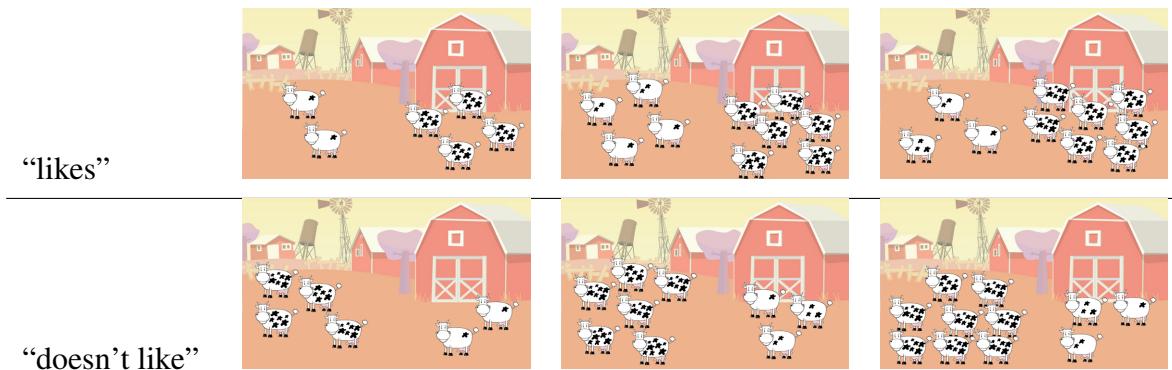


Table 1: **Target sentences:** *The puppet likes the cards where DP are by the barn.*

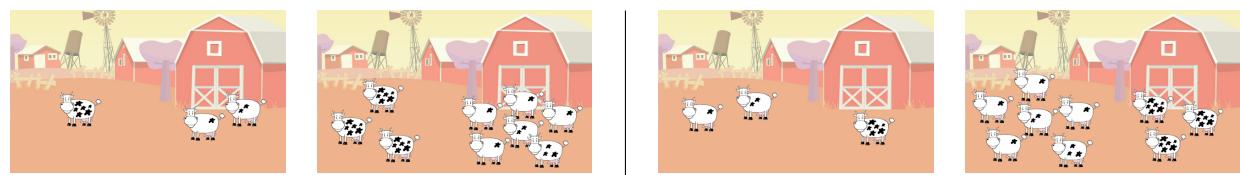
condition	DP	the	partitive
ADJECTIVAL	<i>the gleebest cows</i>	✓	✗
CONFOUNDED	<i>the gleebest of the cows</i>	✓	✓
DETERMINER	<i>gleebest of the cows</i>	✗	✓

Table 2: **Numbers and ratios of cows and spots on training and test cards.** Each H or L represents a unique cow, and indicates whether that cow was high-spotted (H; 6, 7, or 8 spots) or low-spotted (L; 1, 2 or 3 spots). Ratios of cows by the barn to those in the field and spots on the cows by the barn to those in the field for each card are given under column headings of the same name.

Training: Quantity and Quality True				Training: Quantity and Quality False			
barn	field	cows	spots	barn	field	cows	spots
H,H,H,H	L,L	0.50	0.08	L,L	H,H,H,H	0.50	0.08
H,H,H,H,H,H	L,L,L,L	0.67	0.14	L,L,L,L	H,H,H,H,H,H	0.67	0.14
H,H,H,H,H,H,H	L,L,L	0.43	0.14	L,L,L	H,H,H,H,H,H,H	0.43	0.14

Test: Quantity-True, Quality-False				Test: Quantity-False, Quality-True			
barn	field	cows	spots	barn	field	cows	spots
L,L	H	0.50	0.50	H	L,L	0.50	0.50
L,L,L,L,L,L,L	H,H,H	0.50	0.50	H,H,H	L,L,L,L,L,L,L	0.50	0.50
L,L,L	H,H	0.67	0.67	H,H	L,L,L	0.67	0.67
L,L,L,L,L	H,H	0.40	0.40	H,H	L,L,L,L,L	0.40	0.40
L,L,L,L,L	H,H,H	0.60	0.60	H,H,H	L,L,L,L,L	0.60	0.60
L,L,L,L,L,L,L	H,H,H,H	0.57	0.57	H,H,H,H	L,L,L,L,L,L,L	0.57	0.57

Figure 2: Sample unambiguous test cards.



Quantity True, Quality False

Quantity False, Quality True

Table 3: **Predicted neutral (–) versus increased quantity-based responses (+).**

Hypothesis	ADJECTIVAL	CONFOUNDED	DETERMINER
Category as strong cue	–	–	+
Partitive as a strong cue	–	+	+
No bootstrapping	–	–	–

Figure 3: Percent quantity-based responses by condition.

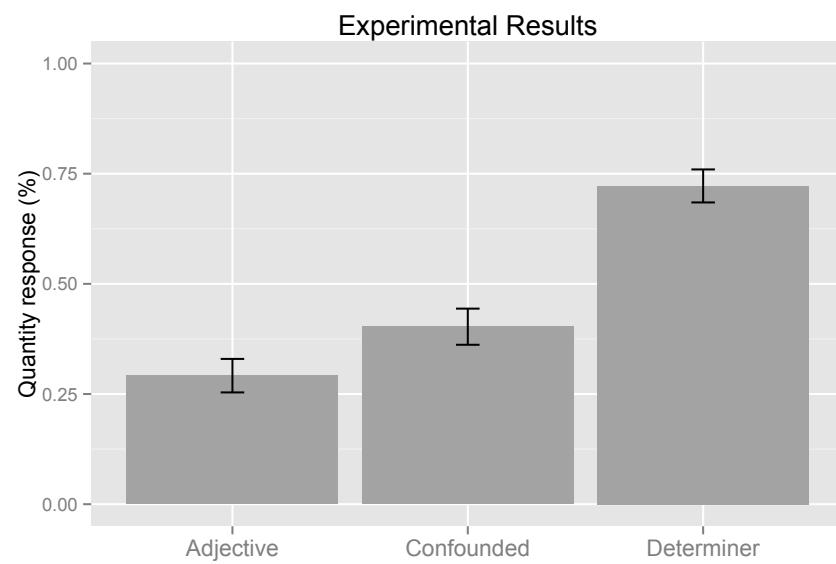
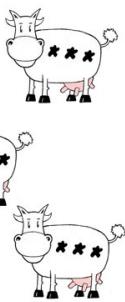
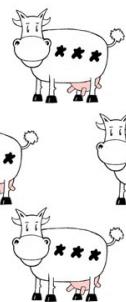
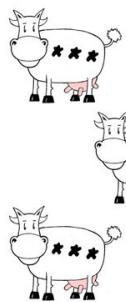
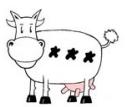
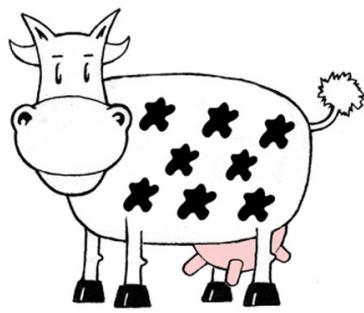
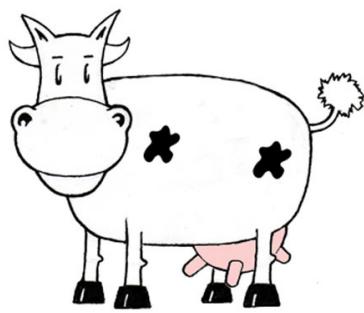


Figure 4: Sample images used in the adult similarity judgment task.



Quantity comparison



Quality comparison

Figure 5: Hierarchical clustering of mean similarity judgments.

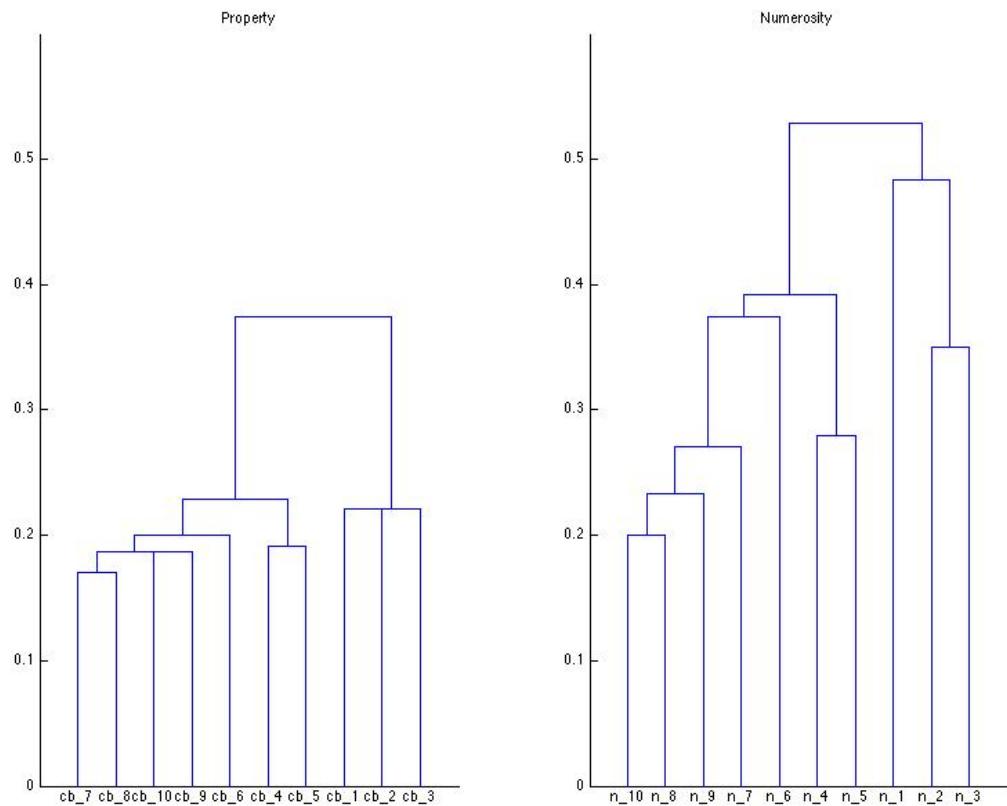


Table 4: **Mean cluster distinctiveness** ($\mathbb{P}(h^{saliency})$)

Quantity	Quality
5.5	4.6

Table 5: Terms and parameters in Model 4 - Perceptual Bias

Term	A	B	C	D
Quantity encoded correctly?	no	yes	no	yes
Quality encoded correctly?	yes	no	no	yes
Parameter	$\alpha = \mathbb{P}(A)$	$\beta = \mathbb{P}(B)$	$\gamma = \mathbb{P}(C)$	$\delta = \mathbb{P}(D)$
Parameter value	$\alpha = \text{free}$	$\beta = \text{free}$	$\gamma = \alpha * \beta$	$\delta = 1 - \alpha - \beta - \gamma$

Table 6: **Approximated confusability values used in simulations.**

α	β
0.2	0.025

Table 7: Role of lexical bias and salience in each model

	Model 1: Lexical Bias	Model 2: Conceptual Bias	Model 3: Lex .& Concep. Bias	Model 4: Perceptual Bias
Lexical bias	prior	–	prior	prior
Salience	–	prior	prior	encoding

Figure 6: Simulations results compared with child data (ADJECTIVAL).

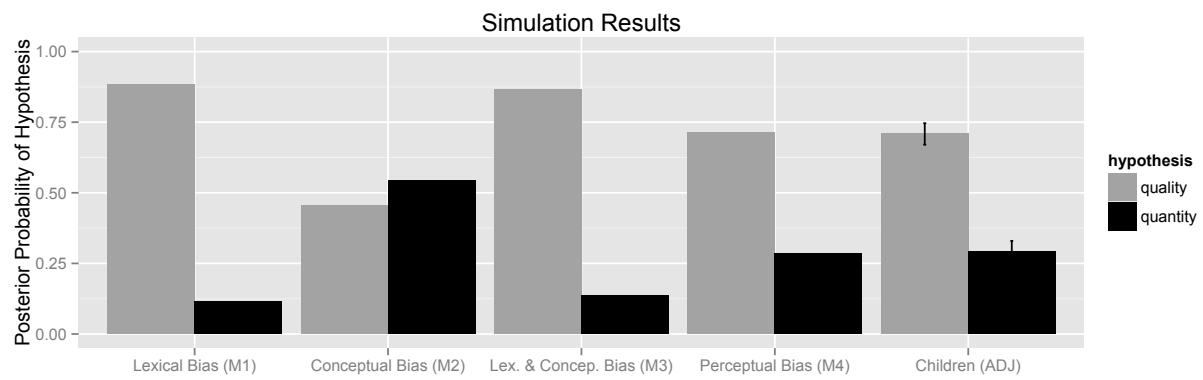


Figure 7: **Similarity judgments across ratios, collapsing quantity- and quality-based judgments.**

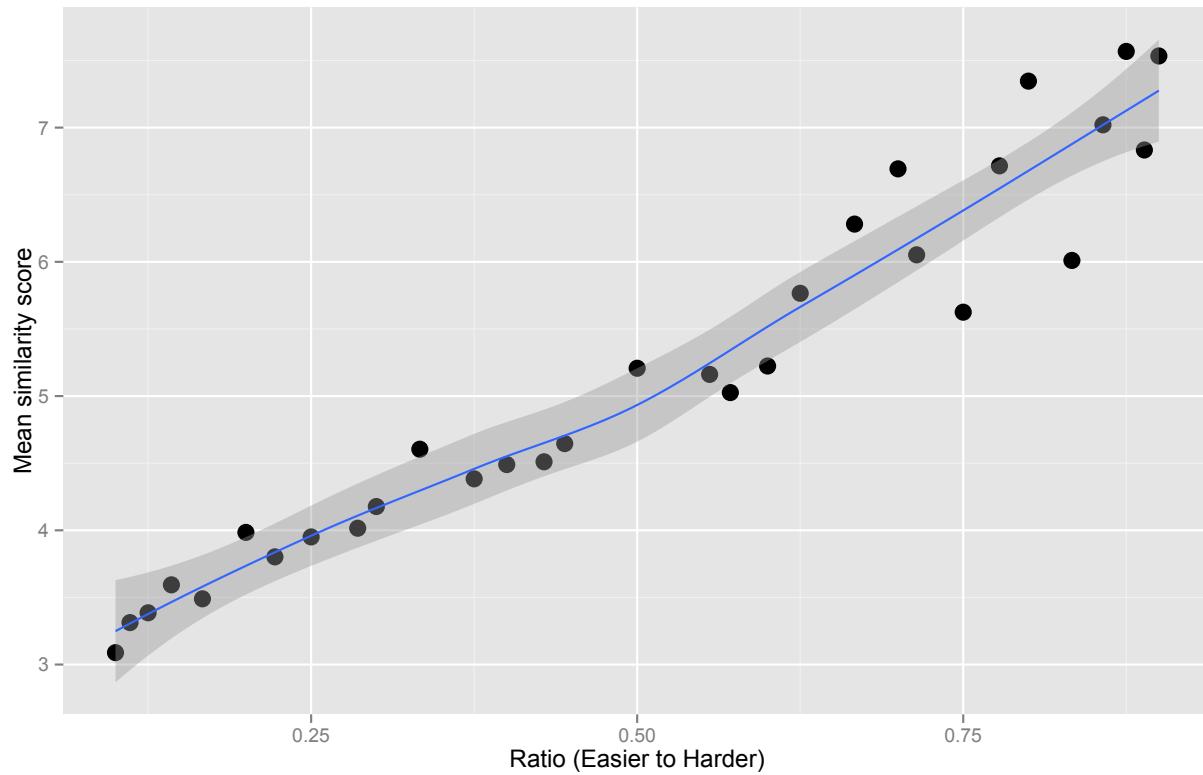


Figure 8: Comparison of similarity judgments across ratios for quantity- and quality-based judgments. Colored numbers represent projected similarity values for ratios used in all six experimental training trials, with red representing quality comparisons and blue representing quantity comparisons.

