Available online at www.sciencedirect.com
ScienceDirect
Cognition 107 (2008) 19–53
COGNITION
www.elsevier.com/locate/COGNIT

Preschool children's use of cues to generic meaning ☆

Andrei Cimpian *, Ellen M. Markman

Department of Psychology, Stanford University, Jordan Hall, Building 420, 450 Serra Mall, Stanford, CA 94305-2130, USA

Received 1 September 2006; revised 4 July 2007; accepted 20 July 2007

Abstract

Sentences that refer to categories – generic sentences (e.g., "Dogs are friendly") – are frequent in speech addressed to young children and constitute an important means of knowledge transmission. However, detecting generic meaning may be challenging for young children, since it requires attention to a multitude of morphosyntactic, semantic, and pragmatic cues. The first three experiments tested whether 3- and 4-year-olds use (a) the immediate linguistic context, (b) their previous knowledge, and (c) the social context to determine whether an utterance with ambiguous scope (e.g., "They are afraid of mice", spoken while pointing to 2 birds) is generic. Four-year-olds were able to take advantage of all the cues provided, but 3-year-olds were sensitive only to the first two. In Experiment 4, we tested the relative strength of linguistic-context cues and previous-knowledge cues by putting them in conflict; in this task, 4-year-olds, but not 3-year-olds, preferred to base their interpretations on the explicit noun phrase cues from the linguistic context. These studies indicate that, from early on, children can use

☆ We are grateful to the children, parents, and teachers at Bing Nursery School. We also thank Heidi Hamilton Kerko, Katie Ostrager, Stephanie Cheng, and Caitlin McLean for their assistance in testing and coding the videotapes. Lera Boroditsky, Joe Robinson, Frances Chen, and our anonymous reviewers provided helpful comments on this research and on the manuscript. We are indebted to Joe Robinson and Ewart Thomas for their help with the statistical analyses. These results were presented at the March 2007 meeting of the Society for Research in Child Development in Boston, MA, USA.

* Corresponding author. Tel.: +1 650 804 2475.
E-mail address: acimpian@psych.stanford.edu (A. Cimpian).

doi:10.1016/j.cognition.2007.07.008

contextual and semantic information to construe sentences as generic, thus taking advantage of the category knowledge conveyed in these sentences.

Keywords: Generics; Language development; Conceptual development; Pronoun comprehension

1. Introduction

It is beyond doubt that, across a wide range of contexts, children's knowledge acquisition relies to a considerable extent on the information provided by others through language (e.g., Harris, 2002; Jaswal, 2004; Koenig, Clément, & Harris, 2004). However, language is particularly influential when the information it expresses is not easily derived through first-hand observation and experience (Harris, 2002). In this paper, we focus on a powerful linguistic means of conveying knowledge: kind-referring *generic sentences* (or *generics*), which are sentences that express generalizations about categories (e.g., "Dogs are friendly"). Generics are likely to be important to children's conceptual development for several reasons. First, there are no comparable non-linguistic means of unambiguously conveying that a property applies to a category as a whole (Gelman, 2003; Gelman, 2004). Pointing to any number of friendly dogs cannot substitute for the generic "Dogs are friendly", as the category *dogs* also covers past, future, or hypothetical exemplars (Gelman, 2004). Furthermore, some generic meanings (e.g., that dogs are widespread) cannot even be illustrated by displaying individual examples (Heyer, 1990): No individual dog can be said to be widespread. Second, generics are particularly robust even in comparison to other linguistic devices. For instance, unlike universally quantified sentences (e.g., "All dogs are friendly"), generics are resistant to counterexamples (e.g., Prasada, 2000): "Dogs are friendly" remains true even after an encounter with a mean dog, but "All dogs are friendly" does not. This feature makes generic sentences ideally suited to express properties that are typical or important for a category but that can nevertheless admit exceptions. Third, generics may scaffold children's own inductive inferences (Cimpian & Markman, 2005; Gelman, Star, & Flukes, 2002; see also Cimpian, Arce, Markman, & Dweck, 2007). Consider, for example, the many possible inferences a child could make after seeing a single friendly dog: Is it just this dog that is friendly, or is it dogs in general? Or maybe pets, or furry things? Since the available evidence is logically compatible with an infinity of inductions (Goodman, 1965; Skyrms, 1986), children may use the presence of generic vs. non-generic sentences to decide whether (and how far) they should generalize. Fourth, generic sentences are quite frequent in speech addressed to young children (e.g., Gelman, Coley, Rosengren, Hartman, & Pappas, 1998; Gelman & Tardif, 1998). Last, it is likely that generic meaning can be expressed in all languages (Behrens, 2005; Carlson & Pelletier, 1995; Gelman, 2004).

Despite their potential role in knowledge acquisition and transmission, generic sentences cannot be identified by a simple rule. On the contrary, determining whether a sentence is generic involves attending to a complex system of grammatical, semantic, and pragmatic cues. In this paper, we explore 3- and 4-year-old children's ability to use various sources of information to resolve this interpretive problem. First, however, we take a closer look at what makes generic sentences so challenging.

1.1. The problem of generic language

Many meanings expressible in language can be conveyed solely through morphosyntactic means. For example, when referring to something that happened in the past, English speakers can either add the suffix *-ed* to a verb or use a verb-specific irregular past tense form. These formal cues are sufficient to express the meaning that the action or event described is completed, although speakers can certainly specify further what they mean with the use of gestures or additional words (e.g., "yesterday" or "last week"). In contrast, no combination of formal grammatical devices (e.g., tense, number) can ensure that an utterance that is generic from the speaker's perspective will actually be interpreted as such. Whether a sentence is understood as generic often depends as much on the extrasentential context and the listener's world knowledge as on the particular morphosyntactic devices used (Declerck, 1986; Declerck, 1987; Declerck, 1991; Gelman, 2004; Gelman & Raman, 2003).

1.1.1. Morphosyntactic cues

There are very few restrictions on the types of noun phrases and verbs that can appear in a kind-referring generic sentence. For example, generic noun phrases can have any of the following forms (Carlson & Pelletier, 1995; Dahl, 1975; Declerck, 1986; Declerck, 1991; Gelman, 2004; Geurts, 1985; Greenberg, 2003; Heyer, 1990; Papafragou, 1996):

(1) Mass nouns: Gas is expensive.
(2) Singular nouns with definite articles: The mouse is a rodent.
(3) Singular nouns with indefinite articles: A car can be very useful.
(4) Bare plural nouns: Tigers are massive.
(5) Singular pronouns: It [a cactus] grows in the desert.
(6) Plural pronouns: They [beavers] build dams.

Definite plural nouns (e.g., "the tigers"), on the other hand, are not acceptable in a kind-referring generic sentence (Declerck, 1991; Gelman & Raman, 2003). Hence, a sentence such as

(7) The tigers are massive.

would be understood as referring to a specific set of tigers rather than the kind.

Despite their compatibility with generic meaning, the noun phrases in (1)–(6) can also appear in non-generic sentences. Consider the following sentences:

(8) Gas trickled out of the engine.
(9) The mouse looked scared.

(10) A car almost hit me.
(11) Tigers escaped from the zoo.
(12) It [a cactus] was very prickly.
(13) They [beavers] chewed through a tree trunk.

Sentences (1)–(6) and (8)–(13) have identical noun phrases, yet the latter are not kind-referring. It is thus apparent that no combination of cues relevant to noun phrases – number and article – can by itself guarantee that a sentence is understood as referring to a kind.

Although their noun phrases are the same, sentences (1)–(6) and (8)–(13) do differ systematically in verb tense. Generic sentences are often in the present tense, but tense alone is also insufficient to guarantee generic meaning – consider

(14) The mouse is in its hole.

Sentence (14) is non-generic, and so are countless other sentences that are in the present tense but refer to a limited set of objects. What's more, some perfectly acceptable generic sentences have verbs in the past or future tenses:

(15) The dog was domesticated many thousands of years ago.
(16) Dogs will always be our trusty companions.

Another verb-related cue that often correlates with generic meaning is aspect: Most generic sentences have verbs in the simple aspect (e.g., [1]–[6]). Again, however, there are many more non-generic than generic sentences in the simple aspect (e.g., [7]–[16]), and acceptable generic sentences in both progressive and perfect aspects do exist:

(17) Gas is getting more and more expensive.
(18) Hurricanes have become extremely violent.

Bowdle and Ward (1995) provide another compelling example of how formal cues can underdetermine interpretation of a sentence with respect to the generic/non-generic distinction. Under most circumstances, the presence of a demonstrative pronoun such as "that" or "those" leads to a non-generic interpretation:

(19) Those iPhones need to be shipped right away.

However, in evaluative contexts, nouns accompanied by demonstrative pronouns can sometimes refer to a kind, as in the following exchange (see Bowdle & Ward, 1995, for other examples):

(20) A: I just bought myself an iPhone.
(21) B: Oh, those iPhones are so cute!

In sum, a learner trying to distinguish between generic and non-generic utterances based on grammatical cues alone is faced with a very difficult task, as (a) there are many combinations of such cues that can express generic meaning and (b) most, if not all, of these combinations are actually open to both interpretations. What helps resolve the ambiguity on any occasion is the linguistic and physical context of the utterance, as well as the listener's general knowledge. This being said, though, there does seem to be quite a bit of variation in the extent to which these extrasentential sources of information need to be invoked. The more prototypical ways of expressing generics – for example, with a bare plural and a present tense verb, as in (4) – may be understood without much supporting context. At the other extreme, generic sentences like (21) seem truly underdetermined by their grammatical form.

1.1.2. Context cues

Where there is ambiguity, generic meaning can often be inferred by considering the linguistic context of an utterance, as in the examples below:

(22) Dogs make good pets. They are loyal and friendly.
(23) These dogs will make good pets for you. They are loyal and friendly.

"They" is understood as referring to the category in (22) and some specific exemplars in (23), based on the content of the first sentence in each example. Experiment 1 in this paper will test children's sensitivity to a simplified version of this linguistic cue. Disambiguating cues are not always linguistic, though – sometimes the physical context of an utterance can clarify its meaning (Gelman & Raman, 2003; Pappas & Gelman, 1998):

(24) They are really good pets. [pointing to a single dog]
(25) They are really good pets. [pointing to two or more dogs]

In (24), the mismatch between the number of objects present and the plural pronoun signals generic meaning; children as young as 3 are sensitive to this pragmatic cue (Gelman & Raman, 2003). In (25), however, the speaker's meaning is more ambiguous, and a non-generic interpretation (i.e., that only the dogs that the speaker is pointing to are good pets) is plausible.

The non-linguistic context can also influence the interpretation of an utterance of ambiguous scope by conveying information about the speaker's knowledge. Consider the following sentence, spoken while pointing to two dogs:

(26) They can't see colors well.

If (26) was uttered by a veterinarian upon completing an examination (and assuming that the addressee has no prior knowledge about dog vision), the sentence would probably be understood as referring to the two dogs pointed to. Having just examined them, the veterinarian has first-hand knowledge of their health and is likely to provide a diagnosis. In contrast, if the same sentence was spoken by a biology teacher pointing to a couple of dogs in a textbook, the preferred interpretation would be generic. The teacher has no prior familiarity with these particular dogs and – especially in the context of the classroom – is more likely to impart information about categories rather than individuals. Children's ability to capitalize on these contextual cues was investigated in Experiment 3.

1.1.3. Knowledge cues

General knowledge of properties and categories can also be of use in determining whether or not a sentence is kind-referring. First, properties vary in the degree to which they are applicable to an entire category. Properties such as *being wet* or *being sick* usually describe temporary states, so sentences like the following are likely to be interpreted as referring to a specific individual:

(27) The cow is sick.

In contrast, biological or behavioral properties are often part of general overhypotheses (Goodman, 1965; Shipley, 1993; Shipley, 2000) and are easily generalizable to an entire category:

(28) The cow is soft-hoofed.

Experiment 2 investigated whether children use the contrast between these types of properties in interpreting an ambiguous sentence, and Experiment 4 tested whether this property-type information is weighted more heavily in children's disambiguation decisions than noun phrase information from the linguistic context.

A similar analysis applies to kinds: Some categories are better entrenched (Goodman, 1965) or familiar than others, and better entrenchment often facilitates generic interpretation (example from Carlson, 1977, cited in Carlson & Pelletier, 1995):

(29) The Coke bottle has a narrow neck.
(30) The green bottle has a narrow neck.

Kind reference in (29) is possible (although the final interpretation will have to depend on the context) partly because *Coke bottle* is a coherent, familiar category. Sentence (30), however, does not allow a generic interpretation.

1.2. Previous evidence on children's use of cues to generic meaning

Given the range of information that has to be integrated in interpreting generic sentences, are children actually able to tell them apart from non-generics? If so, what cues can children use to discriminate between generics and non-generics? Two studies (Gelman et al., 2002; Hollander, Gelman, & Star, 2002) showed that preschool children can use the presence of specific lexical items – namely, "all" and "some"– to avoid generic interpretation. Generics are distinct from universally quantified statements (e.g., ‘‘All birds fly’’), in that they allow exceptions. On the other hand, generics are stronger in their implications than indefinite statements (e.g., ‘‘Some birds can fly’’). To test whether children make these distinctions, Gelman et al. (2002) provided 4-year-olds with a property of a target item in one of these three forms (i.e., generic, universally quantified, or indefinite) and then asked children to choose which of a series of test items (from the same category) also possessed that property. Children selected the most test objects for the universally quantified properties (92%) and fewest test objects for the indefinites (62%), with generics in the middle (75%). Thus, it seems that by the age of 4, children have acquired an understanding of generics as conveying a broad generalization about the members of a category that nevertheless allows for exceptions. This conclusion is also supported by the Hollander et al. (2002) study, which replicated the differences between generics, universally quantified sentences, and indefinites with another task.

Gelman and Raman (2003) investigated children’s sensitivity to one morphosyntactic (Study 1) and one pragmatic (Study 2) cue. In their first study, they presented 2- to 4-year-old children with two atypical examples of a category (e.g., two cats without tails) and asked them a single question about the atypical dimension – either one containing a bare plural noun (e.g., ‘‘Do *cats* have tails?’’) or one containing a plural noun with a definite article (e.g., ‘‘Do *the cats* have tails?’’). The combination of bare plural and present simple tense is a prototypical generic form and should lead children to answer based on the kind as a whole (i.e., ‘‘Yes, they do have tails’’), despite the conflicting context. In contrast, if children are sensitive to the presence of a definite article next to the plural noun, they should answer the second question based on the two atypical examples (i.e., ‘‘No, they don’t have tails’’). All children, including the 2-year-olds, gave specific answers for the non-generic question and category-wide answers for the generic question, thus showing sensitivity to this morphological cue.

In Study 2, Gelman and Raman (2003) focused on a pragmatic cue to generic meaning – using a plural noun phrase in the presence of a single object. In previous studies of mother–child conversations (Gelman et al., 1998; Pappas & Gelman, 1998), mothers often produced sentences using plural nouns (e.g., ‘‘Bats live in caves’’) while pointing to single objects (e.g., one bat). In the presence of such discrepancies, Gelman and Raman argued, listeners assume the utterances refer to kinds (see also Gernsbacher, 1991). In contrast, a generic interpretation is less likely in cases where there is a match between the noun phrase and the number of objects in the context. To test this claim, Gelman and Raman presented children with a mismatch situation: They saw an object that was atypical of its category (e.g., a square balloon) and were then asked a question about it in the plural (e.g., ‘‘What shape are they?’’). If children are sensitive to the pragmatic mismatch, they should interpret the question generically and give a category-wide answer (e.g., ‘‘round’’). The crucial comparison is with a situation where children are asked the same question (e.g., ‘‘What shape are they?’’) but about *two* atypical objects. In this case, the match between the noun phrase and the context should give rise to a non-generic interpretation and a specific answer (e.g., ‘‘square’’). The results showed that the 3- and 4-year-olds (but not the 2-year-olds) were sensitive to the contrast information and gave category-wide answers for the mismatch questions (one object + plural) and specific answers to the match questions (two objects + plural).

1.3. The present studies

The pioneering studies of Gelman and her colleagues have been essential in formulating the problem of generic language and setting up the conceptual framework for a systematic investigation of this topic. Nevertheless, many questions remain about young children's ability to interpret generic sentences. For example, we do not know whether children can use information from adjacent sentences when judging whether an ambiguous sentence is generic (see [22] and [23]). Similarly, there have been no demonstrations of children's ability to bring their knowledge about the world (see [27]–[30]) or about others' mental states (see [26]) to bear on such judgments. The present paper introduces a paradigm for testing children's sensitivity to different types of generic cues and uses it to investigate the three cues mentioned above: the immediate linguistic context (Experiment 1), general semantic knowledge (Experiment 2), and information about the social context and other people's knowledge (Experiment 3). In addition, Experiment 4 sought to determine the relative importance of the linguistic context and children's general knowledge by putting these two cues in conflict.

Briefly, the paradigm used in these experiments relies on the ambiguity inherent in producing a sentence with the plural pronoun "they" as a subject in the presence of two or more exemplars of the same category, as in (26). Recall that a similar context was also set up in the two objects + plural pronoun condition in Gelman and Raman (2003) second study. There, however, the children were shown two atypical exemplars (e.g., two square balloons) and asked precisely about the dimension on which the objects were atypical (e.g., "What shape are they?"). Therefore, the context strongly suggested a non-generic interpretation. In the more general situation where a statement is made about two or more objects using "they", the scope of that statement is often ambiguous between the entire category and just the exemplars present. Going back to (26) ("They can't see colors well"), without knowledge of dog biology or information about the speaker and the context, it is difficult to decide whether it is only those two dogs or dogs in general that have poor color vision. It is this ambiguity that we have capitalized on: In all of the studies reported here, the experimenter showed children a pair of objects of the same kind (e.g., two cats) and provided them both with an ambiguous sentence (e.g., "They like to play with toy cars") and with a disambiguating cue, which varied across experiments. The objects were then removed and children had to report what they learned to a stuffed animal that had not been present when the objects were in view. For someone who has not seen the objects, "they" has no referent, so children needed to select another noun phrase (e.g., "cats", "those cats", "the cats") in order to communicate to the stuffed animal what they had learned. Children in an age group were deemed able to use a certain cue if those who received the generic-biasing cue produced generic noun phrases (by saying, e.g., "Cats like to play with toy cars") more often than those who received the non-generic-biasing cue.

2. Experiment 1

This experiment tested whether children are able to determine the scope of an ambiguous target sentence by referring back to a generic or non-generic noun phrase in a preceding sentence. Specifically, for children in one condition, the target sentence was preceded by "Let me tell you something about Xs", whereas in the other condition the disambiguating sentence was "Let me tell you something about these two Xs". If children can take advantage of the additional information in the context sentence, they should produce more kind-referring generics in the preceding-generic condition than in the preceding-non-generic condition. To find out how children interpret the target sentences in the absence of these linguistic cues, we also included a baseline condition, in which the context sentence was uninformative ("Let me tell you something").

2.1. Method

2.1.1. Participants

Seventy-two children from a university-affiliated preschool participated in this study: 36 3-year-olds ($M = 3$ years 7 months, range = 3 years 1 month to 3 years 11 months) and 36 4-year-olds ($M = 4$ years 6 months, range = 4 years to 5 years 2 months). Six additional 3-year-old children were also tested but not included in the final sample because they were not able to complete the task. Equal numbers of boys and girls participated. Children came from predominantly middle-class families.

2.1.2. Design

Equal numbers of children were randomly assigned to three conditions that differed in whether the context sentence contained no informative noun phrase (the baseline condition), a generic noun phrase (e.g., "cats", "rabbits"; the preceding-generic condition), or a non-generic noun phrase (e.g., "these two cats", "these two rabbits"; the preceding-non-generic condition). The experiment consisted of 6 trials, and the order of the trials was counterbalanced across children.

2.1.3. Materials

We used six 15 cm × 21 cm realistic color pictures, each consisting of two animals of the same kind. For each pair, the experimenter provided a context sentence followed by a property (see Table 1 for the full list). Children subsequently reported what they learned to a small stuffed animal ("Mr. Elephant") approximately 15 cm in height.

These properties were purposely chosen to be ambiguous in scope (e.g., "They are afraid of mice" on the *bird* trial). To validate our intuitions about property ambiguity, we tested 24 undergraduates in a paper-and-pencil version of the baseline task. For each pair of animals, the participants were asked to choose which of two disambiguating sentences, one generic (e.g., "Birds are afraid of mice") and one non-generic (e.g., "These birds are afraid of mice"), "best paraphrased" the target property

Table 1
Items used in Experiment 1

Picture	Property
2 Birds	They are afraid of mice
2 Cats	They like to play with toy cars
2 Dogs	They are afraid of raccoons
2 Fish	They like to hide behind rocks
2 Rabbits	They like to sleep on their side
2 Snakes	They like to eat their food without chewing

(e.g., "They are afraid of mice"). Adults did find the properties to be ambiguous, choosing the generic alternative an average of 3.13 times out of the 6 possible; this average did not differ from chance (3 generic choices), $t(23) = .32$, $p = .750$, $d = .06$. Moreover, this chance-level performance was not simply an artifact of averaging over properties that were disambiguated differently: For 4 of the 6 properties, the percentage of adults who selected the generic alternatives varied between 41.7% and 62.5% and was not different from chance, goodness-of-fit χ^2s$(1, N = 24) \leqslant 1.5$, ps $> .220$.

2.1.4. Procedure

Children were tested individually in a quiet room in their preschool. The experimenter first introduced the study to the children:

> Before we start, let me tell you a little bit about this game. So I'm going to show you some pictures, and we're also going to play with Mr. Elephant. Sometimes Mr. Elephant is going to be with us, on the table, and when he's on the table, he can see what we see and hear what we hear. But sometimes he's going to be down on the floor, and when he's there he can't see or hear anything we're talking about, okay? Ready to look at some pictures?

With the toy out of view, the experimenter then brought out the first pair of animals, pointed to them, and said, for instance, "Now let me tell you something [about birds/about these two birds]. They are afraid of mice". The property varied across trials and was repeated once on each trial to facilitate children's memory. The same type of context sentence used was on all 6 trials.

The experimenter then turned the picture over, put it to the side, and brought the toy onto the table. He then asked the child, "Now can you tell Mr. Elephant what you just learned?" Based on recent evidence (Matthews, Lieven, Theakston, & Tomasello, 2006), we expected that children would be sensitive to the fact that the stuffed animal had not seen the picture and would therefore use nouns rather than the pronoun "they" to refer to the objects depicted. In the Matthews et al. (2006) experiment, both 3- and 4-year-olds used more nouns when the addressee could not see the intended referent than when she could. However, if the children in our experiment did produce an ambiguous sentence in response (e.g., by repeating "They are afraid of mice" or omitting the noun phrase in the subject position), the experimenter asked a clarification question (e.g., "Who is afraid of mice?").[1] Also, if the child forgot the property, the experimenter repeated the entire trial. At the end of the trial, children were thanked and praised for their response, and the toy was put away.

The entire session lasted approximately 10 min and was videotaped.

2.1.5. Coding

Given that children's sentences were always in the simple present tense, we used the type of noun phrase (NP) children produced to code their responses into three categories – generic, non-generic, and ambiguous. Bare plural NPs (e.g., "birds") were coded as generic. NPs that contained a definite article, a demonstrative pronoun, or a numeral – for example, "the birds", "those birds", "two birds", etc. – were coded as non-generic. Singular definite noun phrases (e.g., "the bird") were also included in the non-generic category: Although they could in principle refer to a category, generic uses of this type of NP are extremely rare in speech to young children (e.g., Gelman, 2004; Gelman et al., 1998). Responses were deemed ambiguous if they consisted of (a) both generic and non-generic NPs, produced at different points in the same trial; (b) bare singular NPs (e.g., "bird"); or (c) NPs accompanied by the indefinite article (e.g., "a bird"). The latter were coded as ambiguous because they could refer either to one of the two animals in the picture (an indefinite interpretation) or to the category (a generic interpretation). Ambiguous responses were rare, occurring on only 3.5% of the trials.

Children's responses were coded both online – i.e., by the experimenter running the session – and from the videotapes, by a different experimenter. Videotapes were available for 70 of the 72 children. The two coders agreed on 95.0% of the trials, and discrepancies were resolved by discussion.

2.2. Results and discussion

2.2.1. Data analysis

For each child, we counted the number of generic responses made on the 6 trials. Each generic response received a score of 1, each non-generic response received a score of 0, and each ambiguous response was recorded as 0.5.

2.2.1.1. Non-parametric bootstrap statistics. The response measure was not normally distributed (Shapiro-Wilk test, $z = 2.18$, $p = .015$), with many data points clustered around the ends of the [0, 6] range. Also, the variance in the number of generic responses was not homogeneous across the age × condition cells (Levene's test, $F[5, 66] = 10.71$, $p < .001$). Since these basic assumptions of the ANOVA model were violated, we used non-parametric resampling methods – specifically, bootstrapping – to test our hypotheses (e.g., Chernick, 1999; Efron & Tibshirani, 1986, 1993; Fox, 1997;

[1] The scope of children's first on-task responses was rarely ambiguous in the preceding-generic and preceding-non-generic conditions (10% and 0% of all trials, respectively). However, in the baseline condition, where the context sentence ("Let me tell you something") did not provide the animals' name, children's first responses were ambiguous on 53% of trials.

Lunneborg, 2000; Mooney & Duval, 1993). Unlike parametric tests, the bootstrap procedure computes the sampling distribution of a statistic (t, F, etc.) *empirically*, by treating the original n-participant experimental sample as if it were the population and (1) repeatedly drawing *with replacement* "resamples" or replications of size n from it, (2) calculating the relevant statistic for each such bootstrap resample, and (3) aggregating these values into a bootstrap sampling distribution.

Adapting the method described in Fox (1997, pp. 508–509) and Lunneborg (2000, pp. 486–492), we bootstrapped the Wald χ^2 test statistic, which uses the output from a multiple regression – specifically, the coefficients and their variance–covariance matrix – to test simple and composite linear hypotheses regarding the dependent variable.[2] The Wald χ^2 can thus be used as a substitute both for ANOVA-style F tests and for t tests between individual cells of the design. For each hypothesis to be tested, we generated a bootstrap sampling distribution for the relevant Wald χ^2 by calculating it for each of 10,000 resamples (Chernick, 1999, p. 114); we then determined the p value of the observed Wald χ^2 by consulting this custom sampling distribution, tailored to the characteristics of our sample. We took a further step to ensure that our bootstrap sampling distributions were as accurate as possible: Since the default formula for generating the regression coefficients' variance–covariance matrix – necessary for the Wald statistic – assumes homogeneity of variance (e.g., Fox, 1997, p. 216; Wooldridge, 2003, pp. 790–793), we calculated this matrix by bootstrapping instead. That is, we generated the coefficient variance–covariance matrix for each of our 10,000 resamples by using 500 re-resamples[3] rather than relying on the homogeneity-assuming formula. This conservative "bootstrap within bootstrap" or "nested bootstrap" method (Poi, 2004) was used in all subsequent experiments as well, since their data were similar in nature.

2.2.1.2. Non-parametric effect sizes. Measures of effect size that rely on a standardized mean difference (like Cohen's d) are sensitive to violations of the parametric assumptions (Hedges & Olkin, 1984; Kraemer & Andrews, 1982). We thus report a non-parametric measure of effect size, Cliff's δ (Cliff, 1993; Cliff, 1996), which has been shown to perform well under variance heterogeneity and non-normality (del Rosal, San Luis, & Sánchez-Bruno, 2003; Kromrey, Hogarty, Ferron, Hines, & Hess, 2005). Cliff's δ (henceforward δ) represents the *degree of overlap* between the scores in two samples (say, A and B) and ranges from –1 (if all the scores in A are smaller than all the scores in B) to +1 (if all the scores in A are larger than all the scores in B), with 0 signifying perfect overlap.[4] Conveniently, Vargha and Del-

[2] These analyses were performed using the *bootstrap* command (see Stata Press, 2005, pp. 146–165) in Stata 9.2 (StataCorp, 2007).

[3] Efron and Tibshirani (1986, p. 72) suggest that as few as 100 resamples are actually sufficient to estimate the variability of a test statistic with high accuracy (see also Stata Press, 2005, p. 151). More replications (>1000) are needed when estimating confidence intervals or p values, which are more computationally involved.

[4] δ is easily obtained for any particular pair of samples (say, A with size n and B with size m), by comparing all n scores in A with all m scores in B and calculating $\delta = (\#(A_i > B_j) - \#(B_j > A_i))/(n \times m)$, where $i = 1$ to n, $j = 1$ to m, and # stands for "the number of times".

aney (2000, p. 106) have also calculated δ equivalents of the suggested Cohen's d thresholds (Cohen, 1988) for small ($d = .20$), medium ($d = .50$), and large ($d = .80$) effects: $\delta = \pm.11$, $\delta = \pm.28$, and $\delta = \pm.43$, respectively. (These thresholds should be used with caution, however, because the equivalence holds best for normally-distributed samples with homogeneous variance.)

2.2.1.3. No tests against chance. Our conclusions about children's sensitivity to the cues provided in this and subsequent studies will be based on comparisons of generic responding between conditions. That is, we will test whether children provide more generic responses in the presence of a generic-biasing cue (e.g., "about Xs") than in the absence of that cue or in the presence of a non-generic-biasing cue of the same type (e.g., "about these two Xs"). However, we will not perform comparisons against chance because the level of generic responding expected by chance cannot be determined. On the one hand, there are many more ways to produce non-generic responses (e.g., "the birds", "those birds", "two birds") than generic responses (e.g., only "birds") in our task. On the other, it may be that children find it inherently easier to produce generic responses because they are shorter than the non-generic responses. The different characteristics of generic and non-generic responses thus make it impossible to determine what chance level to test against.

2.2.2. Are children sensitive to the immediate linguistic context?[5]

As predicted, children produced overall more generic responses when the ambiguous sentence was preceded by "Let me tell you something about Xs" ($M = 5.67$ out of 6) than when it was preceded by "Let me tell you something about these two Xs" ($M = 2.10$), Wald $\chi^2 = 47.85$, df = 1,[6] $p < .001$, $\delta = .69$. Moreover, the number of generic responses in the baseline condition ($M = 3.42$) was in the middle (see Fig. 1): lower than in the preceding-generic condition, Wald $\chi^2 = 18.73$, df = 1, $p < .001$, $\delta = -.40$, and higher than in the preceding-non-generic condition, Wald $\chi^2 = 4.12$, df = 1, $p = .049$, $\delta = .28$. Interestingly, the 4-year-olds were overall more likely than the 3-year-olds to produce a generic NP ($M = 4.29$ vs. $M = 3.17$), Wald $\chi^2 = 5.65$, df = 1, $p = .024$, $\delta = .25$.

Moreover, the condition differences varied by age, as indicated by a significant age × condition interaction (see Fig. 1), Wald $\chi^2 = 9.00$, df = 2, $p = .027$: The 3-year-olds gave significantly more generic responses in the preceding-generic condition ($M = 5.83$) than in both the baseline condition ($M = 2.13$), Wald $\chi^2 = 27.90$, df = 1, $p = .006$, $\delta = .71$, and the preceding-non-generic condition ($M = 1.54$), Wald

[5] All the analyses in this and subsequent experiments were also performed with ANOVAs and t tests. The results were quite similar, although the bootstrap tests tended to be more conservative than the parametric tests.

[6] The degrees of freedom for the Wald χ^2 indicate how many restrictions on the regression coefficients are being tested simultaneously. Note that the degrees of freedom we report do not figure into our calculation of the p values: Since we constructed a sampling distribution for each individual test by bootstrapping, we did not need to use the asymptotic χ^2 with κ-degrees-of-freedom distribution to find p.

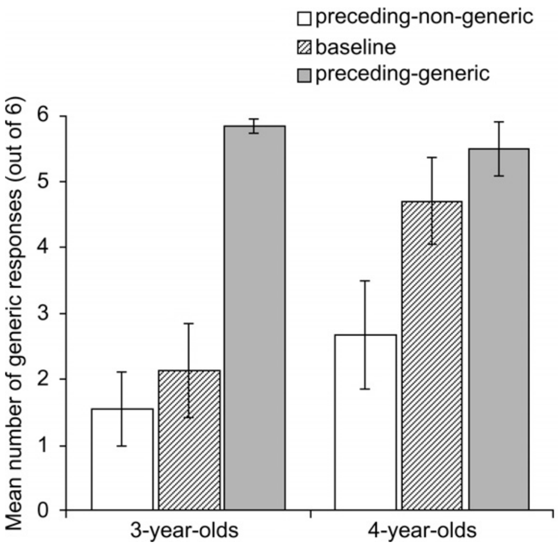

Fig. 1. The mean number of generic responses produced as a function of age and condition in Experiment 1. Error bars represent ± 1 standard error of the mean (*SEM*). The *SEM*s were derived by bootstrapping the mean of each individual cell.

$\chi^2 = 58.35$, df = 1, $p = .004$, $\delta = .90$; the latter two were not different from each other, Wald $\chi^2 = .45$, df = 1, $p = .506$, $\delta = .08$. The 4-year-olds also produced significantly more generic NPs in the preceding-generic condition ($M = 5.50$) than in the preceding-non-generic condition ($M = 2.67$), Wald $\chi^2 = 9.82$, df = 1, $p = .011$, $\delta = .50$. In contrast to the younger children, though, 4-year-olds in the baseline condition produced generic NPs quite frequently ($M = 4.71$); in fact, this condition was not different from the preceding-generic condition, Wald $\chi^2 = 1.12$, df = 1, $p = .318$, $\delta = -.10$, but higher than the preceding-non-generic condition (at the $\alpha = .10$ level), Wald $\chi^2 = 4.24$, df = 1, $p = .058$, $\delta = .40$.

In sum, it appears that 3- and 4-year-olds are able to integrate information across sentences in order to interpret a property ambiguous in scope. That is, when a speaker announces his intention to say something, e.g., "about birds", children are more likely to interpret the subsequent sentence as referring to the entire category *bird* than if the context sentence had been "about these two birds". However, children's behavior in the baseline condition affords a more nuanced interpretation of this result. When given no informative noun phrase cues, 3-year-olds' interpretation of our properties was predominantly non-generic and quite similar to the interpretation they arrived at following "Let me tell you something about these two Xs". In light of these data, claims about the power of the linguistic context for 3-year-olds must be restricted to our generic-biasing cue, "Let me tell you something about Xs". The non-generic-biasing cue did not alter 3-year-olds' default interpretation of the target sentences, and therefore no causal inferences can be made about it from the present data. The situation was reversed for the 4-year-olds: These children's default interpretation of our properties was mostly generic – they actually produced significantly more generic NPs than the 3-year-olds in the baseline condition, Wald $\chi^2 = 7.89$, df = 1, $p = .019$, $\delta = .53$. (This in itself is an interesting finding that we will consider in Section 6.) Thus, our claims about 4-year-olds' sensitivity to linguistic cues are limited to the non-generic-biasing cue, "Let me tell you something about these two Xs", which was used to override children's default (generic) interpretation. On the whole, however, this experiment demonstrates successfully that preschool children are able to use the linguistic context of an utterance in determining its scope.

3. Experiment 2

In this study, we tested whether children can use *knowledge* cues in deciding whether a target sentence is generic or non-generic. To this end, we varied the type of property we presented to the children but kept the linguistic context identical. Two types of properties were used: (a) generalizable, which could plausibly be applied to an entire category (e.g., "They can smell things that are far away" for the *bear* stimulus), and (b) non-generalizable, with a much narrower scope (e.g., "They are sick" for the *bear* stimulus; see, e.g., Gelman, 1988). If children disambiguate these two groups of target sentences differently, we would have evidence that they can use their prior knowledge in determining whether a sentence is generic.

3.1. Method

3.1.1. Participants

Forty-eight children from a university-affiliated preschool participated in this study: 24 3-year-olds ($M = 3$ years 8 months, range = 3 years 2 months to 3 years 11 months) and 24 4-year-olds ($M = 4$ years 9 months, range = 4 years 1 month to 5 years 1 month). Equal numbers of boys and girls were tested. Children came from predominantly middle-class families. Two additional children were also tested but not included in the final sample because they were not able to complete the task (one 4-year-old) or because of experimenter error (one 3-year-old). None of the children had participated in Experiment 1.

3.1.2. Design and materials

Equal numbers of children were randomly assigned to two conditions that differed in whether the properties were generalizable or non-generalizable (see Table 2 for the full list). The context sentence ("Let me tell you something") and the pictures used were identical for the two conditions. As in previous experiments, the pictures were approximately 15 cm × 21 cm in size and depicted two items of the same kind. The categories used were different, though, and consisted of three (new) animal categories and three plant categories (see Table 2). There were six trials, and their order was counterbalanced across children.

To validate our choice of generalizable and non-generalizable properties, we tested 24 undergraduates (the same group as in Experiment 1) on a paper-and-pencil

Table 2
Items used in Experiment 2

Picture	Generalizable property	Non-generalizable property
2 Bears	They can smell things that are far away	They are sick
2 Cows	They have sticky tongues	They are 6 years old
2 Elephants	They can flap their ears to cool off	They are tired
2 Flowers	They have something called "stamen" inside	They come from my grandmother's garden
2 Mushrooms	They grow in dark places	They come from a store near my house
2 Trees	They need something called "carbon" to grow	They are 9 feet tall

version of this task. As before, the adults had to choose between a generic (e.g., "Cows have sticky tongues") and a non-generic (e.g., "These cows have sticky tongues") rephrasing of an ambiguous property (e.g., "They have sticky tongues"). Half of the participants received the generalizable properties, and half received the non-generalizable properties. Adults were able to use their general knowledge to disambiguate the two types of target sentences differently: For the non-generalizable properties, none of the 12 adults chose the generic alternative on any of the trials. In contrast, adults who were tested on generalizable properties selected the generic options for an average of 4.58 of the 6 properties, more often than would be expected by chance (3 generic choices), $t(11) = 3.38$, $p = .006$, $d = .98$.

3.1.3. Procedure

For children in the 4-year-old group, the procedure was almost identical to that of the baseline condition in Experiment 1. The only difference consisted in adding a naming question ("What are these called?") at the beginning of each trial. This question was added because the context sentence ("Let me tell you something") did not contain the name of the category, which is essential for the production of an unambiguous NP. If children did not know what the animals were called or answered incorrectly, they were provided with the answer (e.g., "They're called bears"). As in Experiment 1, the context sentence was followed by a target sentence (whose identity now varied by condition), and then children were asked to report to the stuffed toy what they had learned.

An additional alteration was made to the 3-year-olds' procedure. To lessen the memory load, we changed the experimenter's prompt to, e.g., "Can you tell Mr. Elephant what you just learned? Who can smell things that are far away?" By incorporating the property in the prompt, we no longer required the 3-year-olds to remember and produce the entire target sentence. Rather, they could simply provide a noun phrase that the toy would be able to "understand" (i.e., anything other than "they" or "them").

Video recordings of the experimental sessions were available for all 48 children. Responses were coded into three categories – generic, non-generic, and ambiguous – as in Experiment 1. Ambiguous responses were rare, occurring on only 4.2% of the trials. The inter-coder agreement was 94.4%, and discrepancies were resolved through discussion.

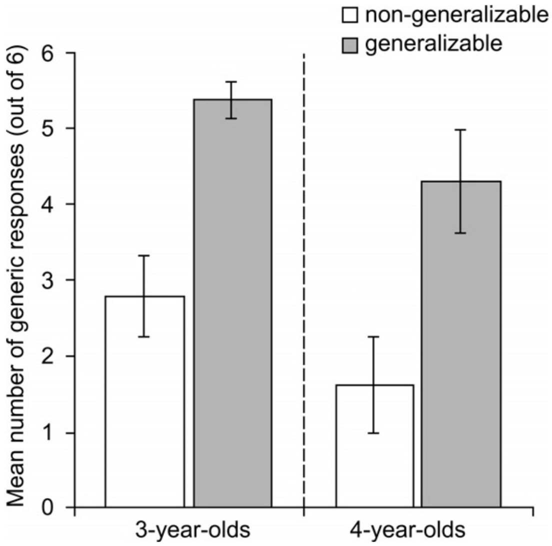

Fig. 2. The mean number of generic responses produced as a function of age and condition in Experiment 2. Error bars represent ± 1 *SEM*. The *SEM*s were derived by bootstrapping the mean of each individual cell. The dashed line indicates that the procedure was different for the two age groups.

3.2. Results and discussion

Because the procedure was not identical for the two ages, we analyzed performance separately within each group. As seen in Fig. 2, the 3-year-olds responded with generic NPs significantly more often when the properties were generalizable ($M = 5.38$ out of 6) than when they were non-generalizable ($M = 2.79$), Wald $\chi^2 = 20.55$, df = 1, $p = .001$, $\delta = .72$. The 4-year-olds showed a similar advantage for generic responses in the generalizable property condition vs. the non-generalizable property condition ($M = 4.29$ vs. $M = 1.63$), Wald $\chi^2 = 7.90$, df = 1, $p = .021$, $\delta = .58$. These results clearly demonstrate that 3- and 4-year-olds children are able to draw on their general knowledge in interpreting a sentence whose scope is ambiguous: When they encountered widely applicable properties, children interpreted the target sentences as generic more often than when the properties were more limited in range.

In contrast to the Experiment 1 baseline (where informative linguistic cues were also absent), the 4-year-olds in this experiment were not more likely than the 3-year-olds to interpret our properties as generic – they actually produced somewhat fewer generic NPs overall than the 3-year-olds ($M = 2.96$ vs. $M = 4.08$), though the difference was not significant, Wald $\chi^2 = 2.76$, df = 1, $p = .110$, $\delta = .22$. Although this discrepancy could be a result of using different types of properties in the two experiments, it is also important to keep in mind that the procedure was not identical for the two age groups in Experiment 2: Unlike the 4-year-olds, who had to formulate a whole sentence in response to the experimenter's question, the 3-year-olds only needed to produce a noun phrase. Thus, it may have been particularly easy for them to just repeat the bare plural NP they produced in response to the naming question (e.g., "bears"), which would be coded as a generic answer. This possible inflation of 3-year-olds' generic responding makes comparisons with Experiment 1 problematic. Regardless, the main point of this experiment stands: Within each age group (and script), children altered their responses depending on the nature of the properties provided, demonstrating sensitivity to this knowledge cue.

4. Experiment 3

In this experiment, we tested whether children can use information about the speaker's knowledge and the social context to disambiguate a target sentence. The intuition here is that the same sentence – say, "They can't see colors well", spoken while pointing to a pair of dogs (see [26]) – may be interpreted differently depending on (a) whether the speaker has first-hand knowledge of the referents and (b) whether the script for the situation calls for a generic or non-generic sentence. For example, if sentence (26) was uttered by a veterinarian who has just examined two dogs, the intended meaning is likely non-generic, both because the vet has knowledge of these particular animals and because an examination is usually followed by a diagnosis (which is, by definition, non-generic). On the other hand, if the same sentence was uttered by a teacher pointing to two dogs in a textbook, its meaning would probably be generic, both because the teacher has no prior acquaintance with the depicted dogs and because teachers usually provide general information about animal kinds. Can preschoolers integrate these types of information into their interpretation of an ambiguous sentence?

4.1. Method

4.1.1. Participants

Forty-eight children from a university-affiliated preschool participated in this study: 24 3-year-olds ($M = 3$ years 8 months, range = 3 years 2 months to 3 years 11 months) and 24 4-year-olds ($M = 4$ years 8 months, range = 4 years to 5 years 5 months). Twenty-three of the children were boys. Children came from predominantly middle-class families. Four additional children were also tested but not included in the final sample because they were not able to complete the task (two 3-year-olds, one 4-year-old) or because of experimenter error (one 4-year-old). None of the children had participated in any of the other experiments.

4.1.2. Design and materials

Equal numbers of children were randomly assigned to either the doctor or the teacher condition. In both conditions, the experimenter read a story about a boy called Eric who goes on a class trip either to a veterinarian's office (the doctor condition) or to a library (the teacher condition). Both stories were printed on letter-size paper and accompanied by colorful pictures, matched as much as possible for content.

The first few pages of the doctor story introduced the main character and briefly described his trip:

> This is Eric. One day, Eric and the other children at his school went to visit a veterinarian's office. A veterinarian is a doctor who takes care of animals. Eric saw many interesting things on his trip, and he wants to show you some of them. Let's take a look!

The teacher story had a similar beginning, except it described a different destination:

> This is Eric. One day, Eric and the other children at his school went to visit the library. A library is a place where there are lots of books. Eric saw many interesting things on his trip, and he wants to show you some of them. Let's take a look!

The following 6 pages consisted of (a) a woman in a white coat with a stethoscope (the doctor) or a woman sitting down (the teacher), (b) two animals of the same kind (e.g., two birds), and (c) the following text: "Eric saw them in the doctor's office. The doctor said to the nurse, [e.g.] 'They have rocks in their tummy'" (the doctor condition) or "Eric saw them in a book. His teacher said to another child, [e.g.] 'They have rocks in their tummy'" (the teacher condition). The doctor or the teacher was present on every page, but the animals and the properties varied (see Table 3 for full list). The order of these 6 pages was counterbalanced across children. The pictures of animals were identical to the ones used in Experiment 1. The properties were chosen so that they were actually true of the category but could also be plausibly interpreted as diagnoses.

To see whether adults would be sensitive to this type of information, we tested 24 undergraduates (12 in each condition) on a paper-and-pencil version of the task. Participants read the introduction and then, on each trial, had to choose which of two sentences, one generic (e.g., "Birds have rocks in their tummy") and one non-generic (e.g., "These birds have rocks in their tummy"), best paraphrased the doctor's or teacher's ambiguous sentence. The adults were not influenced by our manipulation, choosing approximately the same number of generic alternatives in the doctor ($M = 4.67$ out of 6) and teacher ($M = 4.50$) conditions, $t(22) = .24$, $p = .810$, $d = .10$. Recall, however, that all the properties in this study were in fact true of their categories, and adults may have known this: Overall, they chose the category-wide

Table 3
Items used in Experiment 3

Picture	Property
2 Birds	They have rocks in their tummy
2 Cats	They have hair in their ears
2 Dogs	They can't see colors very well
2 Fish	They have a skin bag inside that's filled with air
2 Rabbits	They can't sweat
2 Snakes	They can't hear well

interpretation of the target sentences more often than expected by chance ($M = 4.58$ vs. 3), $t(23) = 4.73$, $p < .001$, $d = .97$. Children's knowledge is probably sparser, though, so they may still be swayed by the contextual information in our stories.

4.1.3. Procedure

Children were tested individually in a quiet room in their preschool. The experimenter introduced children to the stuffed toy and started reading the story. The introductory pages were read without interacting with the child. On each of the six subsequent trials, the experimenter showed children the relevant page, pointed to the two animals on the page and asked "What are these called?", and then read the text, repeating the property once. Next, the page was put away, and children were asked, e.g., "Can you tell Mr. Elephant what the doctor/teacher just said? Who has rocks in their tummy?" (This script was used with both age groups.)

Video recordings of the experimental sessions were available for all 48 children. Responses were coded into three categories – generic, non-generic, and ambiguous – as in Experiment 1. Ambiguous responses accounted for 10.4% of the trials. The inter-coder agreement was 94.4%, and discrepancies were resolved through discussion.

4.2. Results and discussion

Overall, children who heard the teacher story tended to produce more generic responses than children who heard the doctor story ($M = 4.19$ vs. $M = 3.15$ out of 6), although this difference was only significant at the $\alpha = .10$ level, Wald $\chi^2 = 3.13$, df = 1, $p = .099$, $\delta = .25$. There was also a hint of a condition × age interaction, Wald $\chi^2 = 3.13$, df = 1, $p = .099$, but no overall age difference ($M_{3s} = 3.92$ vs. $M_{4s} = 3.42$), Wald $\chi^2 = .74$, df = 1, $p = .420$, $\delta = .16$. Analyzing each age group's responses separately clarified this pattern of results: As seen in Fig. 3, the 4-year-old group produced significantly more generic NPs in the teacher condition ($M = 4.46$) than in the doctor condition ($M = 2.38$), Wald $\chi^2 = 6.63$, df = 1, $p = .029$, $\delta = .58$.[7] On the other hand, the 3-year-olds responded identically in the two conditions (both Ms = 3.92).

Thus, it appears that 4-year-old children can take advantage of subtle cues about the speaker's knowledge and the context of the interaction in determining whether an

[7] We noticed an experimenter effect for the 4-year-olds in the doctor condition, in that only one of the 3 experimenters obtained a really low number of generic responses. Inspection of the tapes did not reveal any consistent differences in the way the 3 experimenters ran the children. Nevertheless, the experimenter who had obtained the highest number of generic responses re-ran the entire doctor condition with another group of 4-year-olds ($N = 12$; mean age = 4 years 7 months) and in fact replicated the overall results from the original doctor condition: On average, children produced only 2.71 generic NPs, not significantly different from the original 2.38, Wald $\chi^2 = .15$, df = 1, $p = .710$, $\delta = .12$. Moreover, there were more generic responses in the original teacher condition ($M = 4.46$) than in the new doctor condition, although the difference was only significant at $\alpha = .10$, Wald $\chi^2 = 3.86$, df = 1, $p = .079$, $\delta = .42$. The teacher vs. doctor difference was also obtained when pooling the two doctor conditions, Wald $\chi^2 = 7.47$, df = 1, $p = .021$, $\delta = .50$. In sum, despite the initial experimenter effect, the low number of generic responses in the doctor condition with 4-year-olds is not specific to a single experimenter.

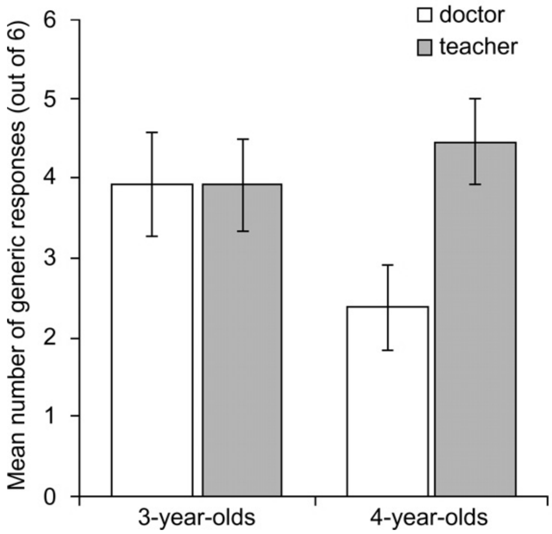

Fig. 3. The mean number of generic responses produced as a function of age and condition in Experiment 3. Error bars represent ± 1 *SEM*. The *SEM*s were derived by bootstrapping the mean of each individual cell.

utterance is generic or not. Sentences spoken by a veterinarian in the presence of two animals were more often interpreted as referring to just those animals than sentences spoken by a teacher who pointed to the same animals in book. Three-year-olds, on the other hand, were not sensitive to these cues.

5. Experiment 4

The interpretation of a sentence as generic is seldom based on the presence of a single cue. On the contrary, listeners routinely need to coordinate multiple sources of information (grammatical, pragmatic, etc.) in order to arrive at an adequate interpretation (e.g., Gelman, 2004). Having multiple cues can sometimes simplify the interpretive problem, especially if all the cues point to the same construal. For example, the sentence "These bears are sick" is less ambiguous than "They are sick" (Experiment 2), owing to the added information conveyed by the noun phrase. On occasion, however, the available cues may point in different directions. For instance, in interpreting a sentence like "This butterfly is very rare" (e.g., Declerck, 1991), a listener would need to resolve the conflict between the demonstrative NP ("this butterfly"), which pulls toward a non-generic reading, and their knowledge of the property (being rare), which is not applicable to a single individual. How do children respond in situations where their world knowledge conflicts with the morphosyntactic information available? Which type of information would they favor? In this study, we pitted linguistic cues in the form of generic NPs in the preceding context (see Experiment 1) against knowledge cues in the form of non-generalizable properties (see Experiment 2). Thus, in the *conflict* condition children heard, e.g., "Let me tell you something about elephants. They are tired". To get a baseline of children's behavior, we also ran a no-conflict condition in which the generic NP was omitted (e.g., "Let me tell you something. They are tired").

5.1. Method

5.1.1. Participants

Forty-eight children from a university-affiliated preschool participated in this study: 24 3-year-olds ($M = 3$ years 5 months, range = 3 years 1 month to 3 years 11 months) and 24 4-year-olds ($M = 4$ years 8 months, range = 4 years 2 months to 5 years 2 months). Equal numbers of boys and girls were tested. Children came from predominantly middle-class families. Three additional 3-year-old children were also tested but not included in the final sample because they were not able to complete the task. None of the children had participated in any of the other experiments.

5.1.2. Design and materials

Equal numbers of children were randomly assigned to two conditions that differed only in their context sentences: In the conflict condition, the context sentence ("Let me tell you something about Xs") contained a generic NP, which was absent in the context sentence for the no-NP condition ("Let me tell you something"). Both conditions used the non-generalizable properties and the pictures from Experiment 2 (see Table 2). There were 6 trials, and their order was counterbalanced across children.

5.1.3. Procedure

On each trial, the experimenter asked children to label the animals and then provided the context and target sentences. The experimenter's subsequent prompt contained the additional memory-facilitating question – e.g., "Can you tell Mr. Elephant what you just learned? Who is tired?"

Video recordings of the experimental sessions were available for all 48 children. Responses were coded into three categories – generic, non-generic, and ambiguous – as in Experiment 1. Ambiguous responses occurred on only 5.9% of the trials. The inter-coder agreement was 96.8%, and discrepancies were resolved through discussion.

5.2. Results and discussion

Children in the conflict condition, where a generic NP was available in the context, produced marginally more generic responses than children in the no-NP condition ($M = 4.77$ vs. $M = 3.67$ out of 6), Wald $\chi^2 = 4.54$, df = 1, $p = .051$, $\delta = .46$. The 3- and 4-year-olds were equally likely to respond with a generic NP when collapsing across conditions ($M_{3s} = 3.90$ vs. $M_{4s} = 4.54$), Wald $\chi^2 = 1.35$, df = 1, $p = .275$, $\delta = .20$. The age × condition interaction was also not significant, Wald $\chi^2 = 2.00$, df = 1, $p = .187$. However, analyzing the two age groups separately revealed that the overall condition difference was driven mostly by the 4-year-olds (see Fig. 4),

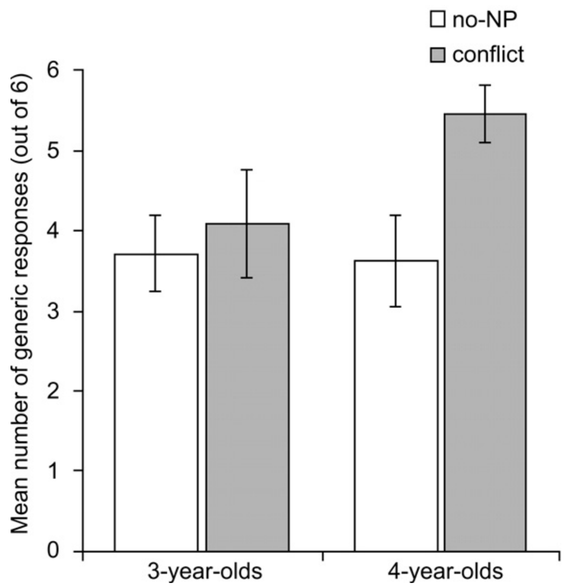

Fig. 4. The mean number of generic responses produced as a function of age and condition in Experiment 4. Error bars represent ± 1 *SEM*. The *SEM*s were derived by bootstrapping the mean of each individual cell.

who responded generically on $M = 5.46$ trials in the conflict condition compared with $M = 3.63$ trials in the no-NP condition, Wald $\chi^2 = 7.41$, df = 1, $p = .020$, $\delta = .65$. The 3-year-olds, on the other hand, were not influenced by the presence of the conflicting generic NP ($M = 4.08$ for conflict vs. $M = 3.71$ for no-NP), Wald $\chi^2 = .23$, df = 1, $p = .643$, $\delta = .25$.

Recall that in Experiment 2, where generalizable and non-generalizable property cues were presented alone (e.g., "They can smell things that are far away" vs. "They are sick"), both the 3- and the 4-year-olds tended to favor non-generic interpretations of the non-generalizable properties – at least relative to the generalizable ones. In this experiment, we tested whether children would depart from this baseline (non-generic) interpretation when provided with a generic-biasing NP in a preceding sentence ("Let me tell you about Xs"). The 4-year-olds produced significantly more generic responses when this linguistic cue was added than in its absence.[8] Thus, they

[8] It is worth noting the difference between 4-year-olds' responses in the no-NP condition in this experiment and their responses in the non-generalizable property condition in Experiment 2. These two conditions were identical with one exception: whether the experimenter asked the additional question (e.g., "Who is sick?"). When they had to respond with a full sentence (Experiment 2), 4-year-olds produced generic responses on 1.63 of the 6 trials. When only an NP was required (this experiment), 4-year-olds produced significantly more generic NPs ($M = 3.63$), Wald $\chi^2 = 6.04$, df = 1, $p = .037$, $\delta = .49$. This increase is consistent with our speculations from Experiment 2: Simplifying the task so that children only need to produce the relevant NP increased the a priori probability of making a generic response (maybe because children were likely to repeat the bare plural they produced in response to the naming question) without necessarily changing how children interpreted the target sentences.

seemed to favor linguistic-context information over property information when the two were at odds. The 3-year-olds' behavior, on the other hand, was unchanged by the addition of the NP cue. This result suggests that 3-year-olds consider their prior knowledge a more reliable cue, though other interpretations are also possible (see Section 6).

The fact that 4-year-olds made more generic responses in the presence of the generic NP cue also clears up an issue from Experiment 1: Since 4-year-olds in that experiment produced as many generic responses when provided with a generic NP in the context sentence as in the absence of any informative NP (see the preceding-generic and baseline conditions in Fig. 1), it was unclear whether they are actually able to use a generic NP cue in interpreting an ambiguous sentence. This experiment shows that they are – and that it may have been the relatively high number of generic responses in Experiment 1's baseline that masked this ability.

6. General discussion

Generic sentences convey useful information about natural and social categories. Yet, identifying which sentences are generic may be a challenge for young children, since the decision cannot be based on a straightforward rule and instead requires integration of multiple sources of information. To test for children's sensitivity to various such sources of information, we introduced a novel method that placed minimal demands on children's verbal production: The experimenter provided sentences with the pronoun "they" as a grammatical subject in the presence of two or more exemplars of a category, which caused the scope of this pronoun to be ambiguous between the particular exemplars and the entire category. Children's task was to produce a noun phrase that disambiguated the experimenter's "they" based on the immediately preceding linguistic context (Experiments 1 and 4), their own knowledge about properties (Experiments 2 and 4), and the social context of the utterances (Experiment 3).

6.1. Summary and further discussion of findings

6.1.1. Experiment 1

In this experiment, which looked at cues from the immediate linguistic context, 3- and 4-year-old children interpreted a target sentence as generic more often if a preceding sentence contained a generic noun phrase ("about Xs") than if it contained a non-generic noun phrase ("about these two Xs"). The baseline condition revealed, however, that children's interpretation of the target sentences in the absence of any noun phrase cues was (a) for the 4-year-olds, similar to their interpretation following the generic NP cues and (b) for the 3-year-olds, similar to their interpretation following the non-generic NP cues. Still, children modulated their interpretation based on at least one of noun phrase cues, demonstrating that they are able to integrate genericity-relevant information across adjacent sentences.

An unexpected result was that the 4-year-olds were overall more likely to produce generic responses than the 3-year-olds, but particularly so in the baseline condition, where the informative NP cues were absent. It is doubtful that children's diverging interpretations were based on their prior knowledge of the properties (e.g., "They are afraid of mice" + 2 birds), especially since even the adults were undecided about most of them. Rather, children's responses may have stemmed from slightly different construals of the task context. For instance, the 4-year-olds may have been more likely to interpret the overall task context – an adult-guided interaction directed at encyclopedia-like animal pictures – as a cue that the experimenter's statements are generic. In general, adults produce a considerable number of generics when talking about pictures of objects – more than in the context of, say, playing with toys (Gelman, Chesnick, & Waxman, 2005). However, the 4-year-olds, who have additional experience with such contexts, may have been particularly sensitive to this situational cue (and thus more likely to assume that the adult will provide generic, category-wide facts). Consistent with this claim, 4-year-olds are also more likely than 3-year-olds to incorrectly recall previously presented non-generic sentences as generic (Gelman & Raman, in press).

6.1.2. Experiment 2

This experiment suggested children can use knowledge cues to disambiguate a target sentence: They interpreted sentences referring to generalizable properties (e.g., properties about insides or sensory abilities) as generic more often than sentences referring to temporary or accidental states (e.g., being tired or sick).

But how do 3- and 4-year-olds figure out which properties are generalizable and which are not? The learning problem here is exacerbated by the fact that the same property may be generalizable with respect to some categories and non-generalizable with respect to others. For example, as our reviewers noted, the non-generalizable properties we used in Experiments 2 and 4 (e.g., being sick) may actually generalize for some categories (e.g., "Patients are sick"). More generally, it seems that different types of properties are relevant to different types of categories (e.g., Murphy, 2004). When reasoning about animal categories, for instance, certain kinds of predicates – about internal organs, habitat, diet, or characteristic behaviors and abilities – are more generalizable than others (e.g., age or height). For artifact kinds, on the other hand, it is properties related to function that are most generalizable, while for material kinds (e.g., glass, wood), it is physical properties such as texture, color, and solubility.

For young children, a likely first step in the process of learning these mappings is to accumulate many specific facts about specific categories (owls see in the dark, cacti grow in the desert, etc.). With this knowledge at their disposal, children may then be in a position to infer more general *overhypotheses* (Goodman, 1965; Shipley, 1993; Shipley, 2000), which map broad classes of properties onto broad classes of kinds. For example, if children know that owls and cats can see in the dark, that moles are blind, etc., they might (implicitly) conclude that different kinds of animals have specific sensory strengths or weaknesses – an overhypothesis. Having this overhypothesis would allow children to generalize novel instances of, say, "sensory strengths or weaknesses" to the relevant categories based on little evidence. Thus, if children were shown a picture of two bears and told, "They can smell things that are far away" (as in Experiment 2), they might readily conclude that bears in general can smell things that are far away. There is in fact evidence that, by preschool age, children have already started to map types of properties onto types of categories (e.g., Gelman & Markman, 1986; Kalish & Gelman, 1992; Springer, 1992). Thus, it is possible that children used overhypothesis-like knowledge to determine whether the ambiguous sentences in Experiment 2 were generic or non-generic.

6.1.3. Experiment 3

In this study, we investigated children's sensitivity to subtle cues about the social context of an utterance and the speaker's knowledge. When the target sentences were spoken by a teacher in the context of a library visit, 4-year-old children were more likely to interpret them as generic than when they were spoken by a veterinarian in her office. However, the 3-year-olds did not modulate their interpretation of the target sentences based on the social context provided in our stories. We speculate that the 3-year-olds may have been unable to use the contextual information due to (a) their limited information-processing abilities, (b) their relative lack of knowledge about the relevant social scripts, or (c) their poor understanding of the causal connection between perceptual/informational access and knowledge.

In this task, children had to remember the information provided in the introduction (regarding the trips to the vet vs. the library) and integrate it with the property provided on every trial to determine the referent of the ambiguous NP. This processing load may have been too high for the 3-year-olds, who often perform more poorly than older children in tasks that rely on working/short-term memory (e.g., Carlson, Moses, & Breton, 2002; Case, Kurland, & Goldberg, 1982; Dempster, 1981) or executive function (e.g., Kirkham, Cruess, & Diamond, 2003; Zelazo, Frye, & Rapus, 1996).

Second, 3-year-olds' understanding of the sequence of events involved in visits to a doctor's office or a library is probably weaker than 4-year-olds'. Research on children's memory for medical examinations (e.g., Ornstein, Baker-Ward, Gordon, & Merritt, 1997; Ornstein et al., 1998) has suggested that children's knowledge of the visit-to-the-doctor script develops considerably over the preschool years, becoming progressively more likely to influence memory and other cognitive processes. In this instance, fewer 3-year-olds may have realized that the vet's utterance probably followed an examination – this point was not explicitly stated in the story.

The other important factor that might be at play in this experiment is 3-year-olds' fragile understanding of the relationship between informational access and knowledge (e.g., Gopnik & Graf, 1988; Wimmer, Hogrefe, & Perner, 1988). In the Wimmer et al. (1988) study, for example, 3-year-olds who saw another child look into a box often denied that that child knew what was in the box. Thus, even if our 3-year-olds did infer that the vet examined the animals, they may not have realized that the examination led to knowledge about those specific animals.

6.1.4. Experiment 4

This study was an attempt to discover what types of information children find more reliable as cues to generic/non-generic meaning. More specifically, we asked whether children would prefer to base their disambiguation decisions on their world knowledge or on explicit NP cues from the linguistic context. Thus, we provided children both with property information (e.g., being sick), which suggested a non-generic reading of the target sentence (Experiment 2), and with linguistic-context information ("Let me tell you about Xs"), which suggested a generic reading (Experiment 1). Four-year-olds provided more generic responses when the non-generalizable properties were preceded by the generic NP cue than when this cue was absent, suggesting that the NP cue was effective in overriding their default interpretation. The 3-year-olds, however, did not produce more generic responses in the presence of this NP cue relative to baseline (the no-NP condition).

It is unlikely that 3-year-old children are simply unable to integrate genericity-relevant information across sentences. In Experiment 1, for example, they successfully incorporated the exact same NP cue ("about Xs") into their interpretation of a target sentence. Rather, 3-year-olds' performance is probably related to the cue conflict we set up. In resolving this conflict, the 4-year-olds relied on the linguistic cue, but the 3-year-olds did not (or could not). We propose two explanations for this result.

It may be that 3-year-olds *were not able* to use the NP cue in this experiment due to the increased information-processing demands of the task. As in Experiment 1 (where 3-year-olds used the NP cue successfully), children had to register the informative noun phrase, hold it in working memory until they heard the target sentence, identify the referent of "they", and generate a response. In this study, however, an extra processing step was required: resolving the conflict between the generic NP cue and the non-generalizable property. This additional demand may have pushed 3-year-olds' working-memory capacity beyond its limit. Another explanation for the developmental difference is that 3- and 4-year-old children simply *assigned different weights* to the conflicting sources of information. That is, 4-year-olds may have been more willing than 3-year-olds to rely on adult-provided information – perhaps because they realized their own knowledge is fallible. Age-related shifts in the relative weight of cues are fairly common across language development (see, e.g., Strohner & Nelson, 1974, on shifts in children's use of semantic and syntactic information in sentence comprehension; cf. Bates et al., 1984; Chapman & Kohn, 1978).

Note that, in this particular task, 3-year-olds' responses may have been more adult-like than those of the 4-year-olds: If adults heard, e.g., "Let me tell you something about elephants. They are tired", they might also ignore the generic noun phrase, dismissing it as a mistake. However, 4-year-olds' strategy of using the explicit linguistic cues provided by adults (rather than their previous knowledge) is probably more beneficial in the long run for children this young. Since their knowledge is often likely to be either insufficient or inaccurate, relying on it consistently – as the 3-year-olds seem to do – may lead to errors of interpretation.

6.2. Relation to the pronoun comprehension literature

Since our task required children to identify the referents of a pronoun (“they”), it may be useful to situate this work in the broader literature on pronoun comprehension. For instance, the developmental work on anaphoric pronouns (e.g., Eisele & Lust, 1996; Lust, Loveland, & Kornet, 1980; Tyler, 1983) is relevant to Experiment 1, where “they” was actually an anaphor referring to a noun phrase from the context sentence (e.g., “birds” vs. “these two birds”). Children’s success in Experiment 1 is consistent with past findings – for example, in Lust et al. (1980) study, 3-year-olds were 70% correct and 4-year-olds 85% correct in interpreting anaphoric sentences such as “*Oscar* bumped the wall when *he* found the penny”. Moreover, as in Experiment 1, children in Lust et al.’s study were able to use NP information from preceding context sentences (e.g., “Now I am going to tell you a little story *about Oscar*”) in their interpretation of anaphoric pronouns.

In Experiments 2 and 3, the ambiguous “they” was not an anaphor, since there was no antecedent NP. To determine its referent, children needed to bring in information about the properties and about the social context, respectively. Thus, our work identifies two sources of information relevant to pronoun interpretation, adding to the previous research in this area – see, for example, the studies on children’s use of the gender of potential referents (Arnold, Brown-Schmidt, Trueswell, & Fagnano, 2004; Arnold, Novick, Brown-Schmidt, & Trueswell, 2001; Scholes, 1981; Tyler, 1983) and their order of mention (Arnold et al., 2004, 2001; Song & Fisher, 2005).

6.3. Experiment 1: An alternative explanation?

In Experiment 1, children produced more generic NPs when the ambiguous sentence was preceded by “Let me tell you something about Xs” rather than “Let me tell you something about these two Xs”. Based on this result, we argued that children are sensitive to NP cues from the immediate linguistic context. However, it is also possible that children were performing this task at a superficial level, merely repeating the experimenter-provided noun phrase (e.g., “birds” or “these two birds”) without truly disambiguating the target sentence. Several lines of evidence argue against this explanation.

First, children’s responses continued to show sensitivity to the cues we provided even when repeating the experimenter’s words was no longer an option (that is, in Experiments 2 and 3). Children’s ability to disambiguate sentences on the basis of subtle cues that did not involve an explicit informative NP makes it less plausible that they had to rely on rote repetition in Experiment 1. Moreover, children as young as 3 are able to assign correct referents to anaphoric pronouns across a variety of tasks that do not require production of the antecedent NP (e.g., true/false tasks, acting-out tasks – see Eisele & Lust, 1996; Lust et al., 1980). In light of this evidence, it is unlikely that children in our experiment needed to resort to repeating the experimenter’s words.

To further assess the repetition hypothesis, we also gathered additional information about children’s behavior from the video recordings. In particular, we recorded

(a) whether a child's verbal response was accompanied by a pointing gesture to the picture of two animals on at least 1 of the 6 trials and (b) what types of noun phrases children produced (e.g., ''those birds'', ''the two birds''). With respect to the pointing data, we found there were more children who pointed to the animals on at least one trial in the preceding-non-generic condition ($n = 8$) than in the preceding-generic condition ($n = 1$), Fisher's exact test $p = 0.010$. This difference held up equally for the 3- and 4-year-olds, Fisher's exact test $p = 1.00$. The increased pointing in the preceding-non-generic condition suggests that children's verbal responses were driven by their understanding of the target sentences rather than by a repetition strategy: Generic sentences report timeless facts that are not tied to any particular exemplars, so pointing is not necessary to clarify their meaning. On the other hand, pointing is often useful in identifying the referents of a non-generic utterance (e.g., ''Those birds are afraid of mice''). Many children in the preceding-non-generic condition made use of this communicative device, which suggests that they understood the ambiguous sentences as referring specifically to the animals in the picture. Importantly, this difference could not have been caused by imitation of the experimenter's points: Recall that the experimenter pointed to the picture in both conditions when introducing the property.

Table 4 presents the data on the specific types of non-generic NPs children produced in the preceding-non-generic and baseline conditions, where the bulk of these responses occurred. We focused on the non-generic responses since children could manifest their understanding of the ambiguous target sentence as a non-generic in a variety of ways – that is, many NPs other than the experimenter-provided ''these two Xs'' were also coded as non-generic. (In contrast, children could respond generically only by producing exactly the same NP as the experimenter – a bare plural.) To assess whether children imitated the experimenter's speech, we can look at the extent to which the non-generic NPs children produced in the preceding-non-generic condition match (a) the experimenter-provided ''these two Xs'' (evidence for imitation) vs. (b) the non-generic NPs they produced in the baseline condition (evidence for independent responding). As seen in Table 4, children in the preceding-non-generic condition rarely repeated the experimenter's NP, using the noun phrase ''these two

Table 4
The types of non-generic noun phrases produced in the preceding-non-generic and baseline conditions in Experiment 1 (in percentages)

Noun phrase	Preceding-non-generic	Baseline
Those Xs	26.4	31.6
The Xs	25.3	49.1
Those two Xs	25.3	—
Two Xs	7.7	—
These two Xs	7.7	—
These Xs	3.3	5.3
The two Xs	2.2	—
The X	2.2	8.8
This X	—	5.3

Xs" on only 7.7% of the trials for which they produced a non-generic response. Moreover, 25.3% of the non-generic responses in this condition consisted of a novel NP, "the Xs"; note that children in the baseline condition also made frequent use of this NP. Even when their responses included a demonstrative pronoun, children in the preceding-non-generic condition used the distal form "those" much more often than the proximal "these", the form the experimenter had used (51.7% vs. 11.0% of non-generic NPs). "Those" (which was also common in the baseline condition) was the more felicitous alternative in the context of our task, since the pictures were at some distance to the side.[9]

Although children did not repeat the experimenter's exact words very often, a comparison of the two conditions in Table 4 does suggest that children's responses were to some extent influenced by what they heard. For example, children in the baseline condition never produced NPs containing the numeral "two", whereas this word occurred in 42.9% of the non-generic NPs in the other condition. (This may be an instance of lexical priming – see, for example, Savage, Lieven, Theakston, & Tomasello, 2003.) On the whole, however, the distribution of children's NPs – along with their pointing behavior, their successful use of subtle cues when imitation was not an option, and their previously documented facility with anaphoric pronouns – does not support the mere-repetition hypothesis. We believe that children were actively making sense of the ambiguous target sentences and telling the stuffed animal – (mostly) in their own words – what they had understood.

6.4. Generics in speech to children

By reviewing the data on the frequency and variety of generic NPs in child-directed speech, we hope to show that the problem of generic language – that is, the lack of a simple rule by which to identify generic meaning – does in fact confront young children in everyday contexts, outside the laboratory.

We first address the issue of the overall frequency of generics: Several studies (Gelman & Tardif, 1998; Gelman, Taylor, & Nguyen, 2004; Gelman et al., 1998; Gelman et al., 2005; Pappas & Gelman, 1998) have now documented the presence of generics in parental speech. For example, Gelman et al. (1998) recorded the number of generics parents produced while reading picture books with their 20- and 35-month-old children. Generics accounted for 3.3% of the total number of utterances and appeared equally often in speech to children of both ages. When factoring in the mean number of utterances per session and the length of the session, this percentage translates into about 30 generics per hour, "a substantial and potentially salient amount of input to children" (Gelman, 2003, p. 171). Moreover, though book-reading does tend to elicit more generics from the parents than other contexts (e.g., play with toys; Gelman & Tardif, 1998; Gelman et al., 2005), this percentage does not

[9] Note that children in Experiments 2–4 used considerably fewer demonstratives in their non-generic answers, relying mainly on definite plural constructions ("the Xs"). Interestingly, the frequency of points to the picture was also lower in these other experiments, suggesting an association between pointing behavior and production of demonstrative NPs (see, for example, Krauss, 1998).

appear to overestimate by too much the generic input parents provide under ordinary circumstances: Gelman, Goetz, Sarnecka and Flukes (in press)[10] analyzed extensive CHILDES (MacWhinney & Snow, 1990) transcripts of 2- to 4-year-old children and their parents and found that generic sentences made up just under 3% of parents' utterances. There were, however, differences across the children's ages, with parents of 2-year-olds producing fewer generics (~1.8% of all utterances) than parents of 3- and 4-year-olds (~3.6% and ~3.3%, respectively).

Parents also use a wide range of morphosyntactic forms to express generics. Gelman et al.'s (1998) monograph included a complete list of the generics produced by the mothers, so we were able to calculate the relative frequencies of different generic constructions: Of the generic NPs in the subject position (which made up the bulk of the generic NPs), 47.4% consisted of bare plurals (e.g., "zebras"), 35.1% consisted of the plural pronoun "they", and 15.8% consisted of indefinite singulars (e.g., "a zebra"); the remaining 1.7% comprised several other types of NPs (e.g., a mass noun ["ice"], a singular pronoun ["he"]). Using their CHILDES sample, Gelman et al. (in press) provided a slightly different estimate of parents' use of generic forms. They coded generic NPs into three main categories – plural NPs (including bare plural nouns and plural pronouns), singular NPs (including indefinite singulars), and mass NPs – and found that these categories accounted for 58%, 23%, and 18% of adults' generic utterances, respectively, with no significant differences between parents of the three age groups. Note that mass generic NPs (e.g., "peanut butter") were encountered much more frequently in the CHILDES sample, probably because food was more likely a topic during these conversations than during the lab sessions.

In spite of the variability, taken together these studies demonstrate that adults produce a substantial number of generics when talking to toddlers and preschoolers. Of particular relevance to our studies is the fact that generic sentences having "they" as a subject are quite common in speech to children (Gelman et al., 1998), suggesting that the ambiguity we exploited (i.e., whether or not "they" is generic) is one that children are often faced with. Moreover, adults use a variety of generic NPs from the beginning; since all of these NPs can also be non-generic, it is likely that even very young children will need to invoke additional sources of information (e.g., semantic, contextual) to distinguish between generic and non-generic statements.

7. Concluding thoughts

Overall, these results illustrate the continuous interaction between the development of children's processing capacities and general knowledge, on the one hand, and the development of their ability to interpret generics, on the other: The more children know and the more efficiently they can process and integrate information, the more likely they are to understand adults' generic and non-generic utterances appropriately. Once absorbed, these sentences add to children's knowledge, which

[10] We thank Susan Gelman for graciously providing us with this article.

in turn enhances their ability to disambiguate future generic and non-generic sentences. This type of spiraling improvement is also observed in the more general mechanisms that have been proposed to account for children's sentence processing (e.g., Fernald, 2005; Fernald, Perfors, & Marchman, 2006; Snedeker & Trueswell, 2004; Trueswell, Sekerina, Hill, & Logrip, 1999). Fernald et al. (2006), for example, describe a ''positive-feedback loop'' involving toddlers' vocabulary knowledge and their efficiency in processing language. Increases in vocabulary size and in the speed of recognizing familiar words lead to more efficient parsing of utterances, which, by freeing up cognitive resources, facilitates the acquisition of new words (and thus further increases in vocabulary). Similar processes also occur later on, as children's on-line language processing becomes progressively more reliant on their ability to incorporate various types of information (semantic, syntactic, distributional, pragmatic/contextual, prosodic) into their interpretation (e.g., Snedeker & Trueswell, 2004). From this point of view, generic sentences are a particularly interesting test case for children's emerging sentence comprehension skills because their meaning is sensitive to so many extrasyntactic factors.

In conclusion, despite the complexities involved in understanding generic sentences, children are reasonably successful at it from early on. In this series of studies, we documented that preschoolers attend to the immediate linguistic context, their previous knowledge, and the social context of an ambiguous utterance in order to determine whether it is generic. This ability to integrate multiple sources of information puts children in a position to benefit from the knowledge expressed in adult speech.

References

Arnold, J. E., Brown-Schmidt, S., Trueswell, J. C., & Fagnano, M. (2004). Children's use of gender and order of mention during pronoun comprehension. In J. C. Trueswell & M. K. Tanenhaus (Eds.), *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions* (pp. 261–282). Cambridge, MA: MIT press.

Arnold, J. E., Novick, J. M., Brown-Schmidt, S., & Trueswell, J. C. (2001). Knowing the difference between girls and boys: The use of gender during on-line pronoun comprehension in young children. In A. H. Do, L. Dominguez, & A. Johansen (Eds.), *Proceedings of the twenty-fifth annual Boston University conference on language development* (pp. 59–68). Somerville, MA: Cascadilla Press.

Bates, E., MacWhinney, B., Caselli, C., Devescovi, A., Natale, F., & Venza, V. (1984). A cross-linguistic study of the development of sentence interpretation strategies. *Child Development, 55*, 341–354.

Behrens, L. (2005). Genericity from a cross-linguistic perspective. *Linguistics, 43*, 275–344.

Bowdle, B., & Ward, G. (1995). Generic demonstratives. In J. Ahlers, L. Bilmes, J. S. Guenter, B. A. Kaiser, & J. Namkung (Eds.), *Proceedings of the twenty-first annual meeting of the Berkeley Linguistics Society* (pp. 32–43). Berkeley, CA: Berkeley Linguistics Society.

Carlson, G. N. (1977). *Reference to kinds in English*. Doctoral dissertation. University of Massachussetts, Amherst.

Carlson, G. N., & Pelletier, F. J. (1995). *The generic book*. Chicago, IL: Chicago University Press.

Carlson, S. M., Moses, L. J., & Breton, C. (2002). How specific is the relation between executive function and theory of mind? Contributions of inhibitory control and working memory. *Infant and Child Development, 11*, 73–92.

Case, R., Kurland, D. M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology, 33*, 386–404.

Chapman, R. S., & Kohn, L. L. (1978). Comprehension strategies in two and three year olds: Animate agents or probable events? *Journal of Speech and Hearing Research, 21*, 746–761.

Chernick, M. R. (1999). *Bootstrap methods: A practitioner's guide*. New York: Wiley.

Cimpian, A., Arce, H. C., Markman, E. M., & Dweck, C. S. (2007). Subtle linguistic cues affect children's motivation. *Psychological Science, 18*, 314–316.

Cimpian, A., & Markman, E. M. (2005, October). *Generics and labels: Two complementary influences of language on inductive reasoning*. Poster presented at the Fourth Biennial Meeting of the Cognitive Development Society, San Diego, CA.

Cliff, N. (1993). Dominance statistics: Ordinal analyses to answer ordinal questions. *Psychological Bulletin, 114*, 494–509.

Cliff, N. (1996). Answering ordinal questions with ordinal data using ordinal statistics. *Multivariate Behavioral Research, 31*, 331–350.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Dahl, Ö. (1975). On generics. In E. L. Keenan (Ed.), *Formal semantics of natural language: Papers from a colloquium sponsored by the King's College Research Centre Cambridge* (pp. 99–111). Cambridge: Cambridge University Press.

Declerck, R. (1986). The manifold interpretations of generic sentences. *Lingua, 68*, 149–188.

Declerck, R. (1987). A puzzle about generics. *Folia Linguistica, 21*, 143–153.

Declerck, R. (1991). The origins of genericity. *Linguistics, 29*, 79–102.

del Rosal, A. B., San Luis, C., & Sánchez-Bruno, A. (2003). Dominance statistics: A simulation study on the d statistic. *Quality & Quantity, 37*, 303–316.

Dempster, F. N. (1981). Memory span: Sources of individual and developmental differences. *Psychological Bulletin, 89*, 63–100.

Efron, B., & Tibshirani, R. J. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science, 1*, 54–75.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall, Inc..

Eisele, J., & Lust, B. (1996). Knowledge about pronouns: A developmental study using a truth-value judgment task. *Child Development, 67*, 3086–3100.

Fernald, A. (2005, April). Young children's inferential use of verb information in learning new object words. Paper presented at the Biennial Meeting of the Society for Research in Child Development, Atlanta, GA.

Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology, 42*, 98–116.

Fox, J. (1997). *Applied regression analysis, linear models, and related methods*. Thousand Oaks, CA: Sage Publications, Inc.

Gelman, S. A. (1988). The development of induction within natural kind and artifact categories. *Cognitive Psychology, 20*, 65–95.

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. London: Oxford University Press.

Gelman, S. A. (2004). Learning words for kinds: Generic noun phrases in acquisition. In D. G. Hall & S. R. Waxman (Eds.), *Weaving a lexicon* (pp. 445–484). Cambridge MA: MIT Press.

Gelman, S. A., Chesnick, R. J., & Waxman, S. R. (2005). Mother–child conversations about pictures and objects: Referring to categories and individuals. *Child Development, 76*, 1129–1143.

Gelman, S. A., Coley, J. D., Rosengren, K. R., Hartman, E. E., & Pappas, A. S. (1998). Beyond labeling: The role of parental input in the acquisition of richly-structured categories. *Monographs of the Society for Research in Child Development*, *63*, Serial No. 253.

Gelman, S. A., Goetz, P. J., Sarnecka, B. W., & Flukes, J. (in press). Generic language in parent–child conversations. *Language Learning and Development*.

Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition, 23*, 183–209.

Gelman, S. A., & Raman, L. (2003). Preschool children use linguistic form class and pragmatic cues to interpret generics. *Child Development, 74*, 308–325.

Gelman, S.A. & Raman, L. (in press). This cat has nine lives? Children's memory for genericity in language. *Developmental Psychology*.

Gelman, S. A., Star, J. R., & Flukes, J. E. (2002). Children's use of generics in inductive inferences. *Journal of Cognition and Development, 3*, 179–199.

Gelman, S. A., & Tardif, T. (1998). A cross-linguistic comparison of generic noun phrases in English and Mandarin. *Cognition, 66*, 215–248.

Gelman, S. A., Taylor, M. G., & Nguyen, S. P. (2004). Mother–child conversations about gender: Understanding the acquisition of essentialist beliefs. *Monographs of the Society for Research in Child Development*, *69*, Serial No. 275.

Gernsbacher, M. A. (1991). Comprehending conceptual anaphors. *Language and Cognitive Processes, 6*, 81–105.

Geurts, B. (1985). Generics. *Journal of Semantics, 4*, 247–255.

Goodman, N. (1965). *Fact, fiction, and forecast*. Indianapolis, IN: Bobbs-Merril Company.

Gopnik, A., & Graf, P. (1988). Knowing how you know: Young children's ability to identify and remember the sources of their beliefs. *Child Development, 59*, 1366–1371.

Greenberg, Y. (2003). *Manifestations of genericity*. New York: Routledge.

Harris, P. L. (2002). What do children learn from testimony? In P. Carruthers, M. Siegal, & S. Stich (Eds.), *The cognitive bases of science* (pp. 316–334). Cambridge: Cambridge University Press.

Hedges, L. V., & Olkin, I. (1984). Nonparametric estimators of effect size in meta-analysis. *Psychological Bulletin, 96*, 573–580.

Heyer, G. (1990). Semantics and knowledge representation in the analysis of generic descriptions. *Journal of Semantics, 7*, 93–110.

Hollander, M. A., Gelman, S. A., & Star, J. (2002). Children's interpretation of generic noun phrases. *Developmental Psychology, 38*, 883–894.

Jaswal, V. K. (2004). Don't believe everything you hear: Preschoolers' sensitivity to speaker intent in category induction. *Child Development, 75*, 1871–1885.

Kalish, C. W., & Gelman, S. A. (1992). On wooden pillows: Multiple classification and children's category-based inductions. *Child Development, 63*, 1536–1557.

Kirkham, N. Z., Cruess, L. M., & Diamond, A. (2003). Helping children apply their knowledge to their behavior on a dimension-switching task. *Developmental Science, 6*, 449–467.

Koenig, M. A., Clément, F., & Harris, P. L. (2004). Trust in testimony: Children's use of true and false statements. *Psychological Science, 15*, 694–698.

Kraemer, H. C., & Andrews, G. (1982). A nonparametric technique for meta-analysis effect size calculation. *Psychological Bulletin, 91*, 404–412.

Krauss, R. M. (1998). Why do we gesture when we speak? *Current Directions in Psychological Science, 7*, 54–60.

Kromrey, J. D., Hogarty, K. Y., Ferron, J. M., Hines, C. V., & Hess, M. R. (2005, August). *Robustness in meta-analysis: An empirical comparison of point and interval estimates of standardized mean differences and Cliff's delta*. Paper presented at the Joint Statistical Meetings, Minneapolis, MN.

Lunneborg, C. E. (2000). *Data analysis by resampling: Concepts and applications*. Pacific Grove, CA: Duxbury Press.

Lust, B., Loveland, K., & Kornet, R. (1980). The development of anaphora in first language: Syntactic and pragmatic constraints. *Linguistic Analysis, 6*, 359–391.

MacWhinney, B., & Snow, C. (1990). The child language data exchange system. *Journal of Child Language, 17*, 271–296.

Matthews, D., Lieven, E., Theakston, A., & Tomasello, M. (2006). The effect of perceptual availability and prior discourse on young children's use of referring expressions. *Applied Psycholinguistics, 27*, 403–422.

Mooney, C. Z., & Duval, R. D. (1993). *Bootstrapping: A nonparametric approach to statistical inference*. Newbury Park, CA: Sage Publications, Inc.

Murphy, G. L. (2004). *The big book of concepts*. Cambridge, MA: MIT Press.

Ornstein, P. A., Baker-Ward, L., Gordon, B. N., & Merritt, K. A. (1997). Children's memory for medical experiences: Implications for testimony. *Applied Cognitive Psychology, 11*, S87–S104.

Ornstein, P. A., Merritt, K. A., Baker-Ward, L., Furtado, E., Gordon, B. E., & Principe, G. (1998). Children's knowledge, expectation, and long-term retention. *Applied Cognitive Psychology, 12*, 387–405.

Papafragou, A. (1996). On generics. *UCL Working Papers in Linguistics, 8*, 165–198.

Pappas, A., & Gelman, S. A. (1998). Generic noun phrases in mother–child conversations. *Journal of Child Language, 25*, 19–33.

Poi, B. P. (2004). From the help desk: Some bootstrapping techniques. *The Stata Journal, 4*, 312–328.

Prasada, S. (2000). Acquiring generic knowledge. *Trends in Cognitive Sciences, 4*, 66–72.

Savage, C., Lieven, E., Theakston, A., & Tomasello, M. (2003). Testing the abstractness of children's linguistic representations: Lexical and structural priming of syntactic constructions in young children. *Developmental Science, 6*, 557–567.

Scholes, R. J. (1981). Developmental comprehension of the third person personal pronouns in English. *Language and Speech, 24*, 91–98.

Shipley, E. F. (1993). Categories, hierarchies, and induction. In D. L. Medin (Ed.). *The psychology of learning and motivation. Advances in research and theory* (Vol. 30, pp. 265–301). San Diego CA: Academic Press.

Shipley, E. F. (2000). Children's categorization of objects: The relevance of behavior, surface appearance, and insides. In B. Landau, J. Sabini, J. Jonides, & E. Newport (Eds.), *Perception cognition and language: Essays in honor of Henry and Lila Gleitman* (pp. 69–85). Cambridge, MA: MIT Press.

Skyrms, B. (1986). *Choice and chance: An introduction to inductive logic* (3rd ed.). Belmont, CA: Wadsworth Publishing Company.

Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology, 49*, 238–299.

Song, H., & Fisher, C. (2005). Who's "she"? Discourse prominence influences preschoolers' comprehension of pronouns. *Journal of Memory and Language, 52*, 29–57.

Springer, K. (1992). Children's awareness of the biological implications of kinship. *Child Development, 63*, 950–959.

Stata Press (2005). *Stata base reference manual: Release 9* (Vol. 1). College Station, TX: Stata Press.

StataCorp (2007). *Stata statistical software: Release 9.2*. College Station, TX: StataCorp LP.

Strohner, H., & Nelson, K. E. (1974). The young child's development of sentence comprehension: Influence of event probability, nonverbal context, syntactic form, and strategies. *Child Development, 45*, 567–576.

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

Tyler, L. K. (1983). The development of discourse mapping processes: The on-line interpretation of anaphoric expressions. *Cognition, 13*, 309–341.

Vargha, A., & Delaney, H. D. (2000). A critique and improvement of the CL Common Language effect size statistics of McGraw and Wong. *Journal of Educational and Behavioral Statistics, 25*, 101–132.

Wimmer, H., Hogrefe, G. J., & Perner, J. (1988). Children's understanding of informational access as source of knowledge. *Child Development, 59*, 386–396.

Wooldridge, J. M. (2003). *Introductory econometrics: A modern approach* (2nd ed.). Mason, OH: Thomson South-Western.

Zelazo, P. D., Frye, D., & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development, 11*, 37–63.