Children’s Interpretation of Generic Noun Phrases

Michelle A. Hollander and Susan A. Gelman
University of Michigan

Jon Star
Michigan State University

Generic utterances (e.g., “Cows say ‘moo’”) have 2 distinctive semantic properties: (a) Generics are generally true, unlike indefinites (e.g., “Bears live in caves” is generic; “I saw some bears in the cave” is indefinite), and (b) generics need not be true of all category members, unlike universal quantifiers (e.g., all, every, each). This article examined whether preschool children and adults appreciate both these features, using a comprehension task (Study 1) and an elicited production task (Study 2). In both tasks, 4-year-old children—like adults—treated generics as distinct from both indefinites (“some”) and universal quantifiers (“all”). In contrast, 3-year-olds did not differentiate among generics, “all,” and “some.” By 4 years of age, generics emerge as a distinct semantic device.

In the present set of studies we examined whether children appreciate these features. There are several possibilities for what generics may at first mean to young children: (a) Generics may receive no consistent interpretation, a possibility that would fit with children’s general difficulty understanding quantifiers such as “all” and “some” (Inhelder & Piaget, 1964). A further potential source of difficulty for children is the complexity in form–function mappings for generic expressions. Not only is the bare plural form (e.g., dogs) used for both generic and nongeneric utterances, but also generics can be expressed in several forms, including bare plurals (e.g., “Tigers are ferocious”), definite singulars (e.g., “The tiger is a ferocious beast”), and indefinite singulars (e.g., “A tiger is a ferocious beast”). (b) Children may at first treat generics as equivalent to “some,” because the same linguistic form (bare plural) is used for both generics and indefinites. (c) Children may at first treat generics as equivalent to “all,” because both express properties that are generally true and children tend to have difficulty learning atypical exemplars (Anglin, 1977; Mervis & Pani, 1980; Nelson, Rescorla, Gruendel, & Benedict, 1978). For example, a child who hears “dogs bark” may be unable to come up with counterexamples and so may interpret the utterance as equivalent to “all dogs bark.”

The question of what generics mean to young children is important for both theoretical and practical reasons. From a theoretical standpoint, children’s capacity to reason about generic noun phrases tells us generally about their capacity to construct generic knowledge. The latter is particularly challenging given the inductive problem encountered when reasoning about generics (Gelman, in press; Prasada, 2000). No amount of personal experience or direct contact can give us access to kinds in their entirety. For example, although one can show a child one (specific) dog, one cannot show a child the generic class of dogs. Likewise, one can never demonstrate, with actual exemplars, photos, or drawings, the distinction between a kind (rabbits) and a plurality of instances (some rabbits). As Waxman (1999) noted, members of object categories are distinct, and often disparate, individuals that tend to appear at different times and places. . . . it would be logically impossible for caretakers to assemble together all mem-

Many languages, including English, distinguish generic utterances (e.g., “Tigers are ferocious”) from nongeneric utterances (e.g., “Those tigers are ferocious”). Generics are an important sort of expression: They refer to kinds rather than individuals (Carlson & Pelletier, 1995), they convey propositions that are broad in scope and relatively essential (Prasada, 2000), they express knowledge that is important for category-based inferences (Gelman, Star, & Flukes, 2002), and they are used frequently in natural speech (Gelman, Coley, Rosengren, Hartman, & Pappas, 1998; Gelman & Tardif, 1998). Recent evidence also suggests that children have begun to distinguish generic from nongeneric utterances in the preschool years (Gelman & Raman, in press; Gelman et al., 2002; Pappas & Gelman, 1998). Despite the centrality of generic language, and despite evidence that generics emerge early in childhood, little is known regarding what these expressions mean to children early in development (Prasada, 2000).

There are two important semantic features of generics: (a) Generics are generally true, and so are distinguished from indefinites (e.g., “Bats live in caves” is generic; “I saw bats in the cave” is indefinite). The distinction between a generic reading and an indefinite reading is particularly critical because the same form of the noun phrase can be used for both (e.g., bats in both examples above). (b) Generics need not be true of all members, and so are distinguished from universal quantifiers (e.g., all, every, each).
A final reason that one cannot point to all instances of a kind is that kinds include both existing and nonexistent instances, including counterfactual instances that one would never encounter.

From a practical standpoint, the interpretation children give to generics has important implications for the kinds of inferences they make. Given how readily adults produce generic noun phrases in their child-directed speech, children are hearing generics on a daily basis. By learning more about how children construe such utterances, we can better understand the linguistic and conceptual input contributing to children’s growing knowledge base.

Most past research examining generic language has focused on the contexts in which parents (e.g., Gelman et al., 1998; Gelman & Tardif, 1998) or children (Pappas & Gelman, 1998) produce generics. Such studies consistently indicate that generics are produced in contexts that systematically differ from the contexts in which nongenerics are produced. For example, generics are especially frequent in talk about animals, whereas nongenerics are not. These results imply that generics are likely to differ in meaning from nongenerics, but it does not tell us how generics and nongenerics differ from one another.

More direct evidence concerning children’s understanding of generic noun phrases establishes a semantic distinction between generics and nongenerics for preschool children (Gelman & Raman, in press). Specifically, for preschool children, generics are more likely to refer to an abstract kind and less likely to refer to the present context, as compared to nongenerics. For example, when shown a picture of two penguins and asked “Do birds fly?” (generic), young children typically report “yes” despite the fact that the birds in the picture don’t fly. In other words, they interpret the generic noun phrase as referring to birds in general. In contrast, when shown a picture of two penguins and asked “Do the birds fly?” (nongeneric), young children typically report “no,” interpreting the noun phrase as referring to the birds in the picture. In that work, however, no information was gathered concerning how generics compare to “all” and “some” in their implications. Thus, although young children honor a semantic distinction between generics and nongenerics, we do not know precisely what form that distinction takes.

We are aware of only one study that has compared generics directly to “all” and “some” expressions (Gelman et al., 2002). In that work, preschool children and adults learned novel properties in either generic form (e.g., “Bears like to eat ants”), “all” form (e.g., “All bears like to eat ants”), or “some” form (e.g., “Some bears like to eat ants”) and then were tested on how far they generalized the new properties to other category exemplars (in this example, to other bears). Results indicated that both 4-year-olds and adults drew the largest number of category-based inferences from “all” statements and the smallest number of category-based inferences from “some” statements, with generic statements intermediate between the two. However, that study did not look directly at what generics mean to children and adults, instead gathering only indirect evidence by examining their implications for inductive generalizations of novel properties. The study was also limited in focusing on a single child age group (4-year-olds).

In this investigation, we directly studied what generics mean to children and adults. We used both a comprehension task (Study 1) and an elicited production task (Study 2) to provide converging evidence on this issue. The studies focused directly on children’s semantic interpretation by comparing children’s understanding of generics with their understanding of “all” and “some.” The studies were also novel in two additional respects: We compared 4-year-olds to a younger age group, 3-year-olds (in Study 1), and we included an examination of the content of generics for children and adults (in Study 2). Altogether, the studies provide new evidence concerning the development of understanding of generics in early childhood.

Study 1: Comprehension of Generics

In Study 1, we focused directly on what generics mean to young children by using a comprehension task. As noted above, for adults, generics are distinctive in implying broad category scope (e.g., “Birds fly” is generally true of birds) yet allowing for exceptions (e.g., penguins). Thus, generics are distinct from both “all” (e.g., “All birds fly”) and “some” (e.g., “Some birds fly”). We conducted an experiment to test whether preschool children appreciate this. The study was modeled after an experiment conducted by Smith (1980) that focused on children’s interpretation of “all” and “some.” In Smith’s (1980) study, children ages 4 years 1 month to 7 years 6 months received a series of questions regarding properties of categories. One third of the properties were true of all members of the category in question (what we call wide-scope properties), one third were true of some members of the category (narrow-scope properties), and one third were true of no members of the category (irrelevant properties). Children were asked about each category–property pairing with either the word all or the word some (e.g., “Do all people have blonde hair?” vs. “Do some people have blonde hair?”). Smith’s (1980) results indicated that even 4-year-olds appropriately distinguished “all” and “some” under favorable presentation conditions (i.e., the first half of the first block of questions).

We predicted that if adults were given the same task with questions presented in generic form, they would treat generics as partly like “all” and partly like “some.” In particular, we predicted that adults would accept both wide-scope properties and (to a lesser extent) narrow-scope properties as true in generic form. These predictions follow rather straightforwardly from the semantic analyses provided earlier. Nonetheless, they would constitute novel empirical demonstrations given the lack of evidence regarding what generics mean to lay adults (for an exception, see Gelman & Tardif, 1998, Study 3). Of primary interest was children’s understanding, relative to these predictions with adults.

Method

Participants. Three-year-old children (N = 18; age range = 3 years 0 months to 3 years 11 months; mean age = 3 years 6 months), 4-year-old children (N = 18; age range = 4 years 0 months to 4 years 10 months; mean age = 4 years 5 months), and adults (N = 36) participated. The adults were undergraduates at a large, midwestern university enrolled in a course in developmental psychology; the children attended local preschools in the same town as the university.

Items. Participants each received three blocks of questions (generic, “all,” and “some”), in one of three orders (all–some–generic, some–generic–all, or generic–all–some). For 4-year-olds and adults, each block consisted of 12 questions: 4 concerning wide-scope properties (e.g., “Are
fires hot?), 4 concerning narrow-scope properties (e.g., “Do girls have curly hair?”), and 4 concerning irrelevant properties (e.g., “Do fish have brains.” The majority of these questions were derived from Smith (1980). Three-year-olds received half as many items as the 4-year-olds and adults. See the Appendix for the full list of properties. Each property was rotated through each of the three wording conditions so that the specific content was not confounded with a particular condition (e.g., across children, a given question would be “Are fires hot?”, “Are all fires hot?” or “Are some fires hot?”).

We had three reasons for presenting the items in blocks rather than intermixing the three kinds of wording over the course of the experiment. First, one benefit of blocking trials is that it minimizes carryover effects. If a question of Type A sets a particular context and a question of Type B sets a different context, then there is concern that a child may have difficulty switching from one context to another context repeatedly throughout the course of an experimental session (see Zelazo & Frye, 1998, for evidence that 3-year-old children have difficulty switching flexibly between rules). By blocking the items, the amount of switching is minimized and a more sensitive test is thus provided. Second, we wished to break up the task to maintain children’s interest. The design of Study 1 required many trials per participant (as noted above, 36 at age 4 and 18 at age 3). By chunking the task into three parts, we were able to maintain children’s interest throughout the task. Third, Smith (1980), whose studies formed the basis of this study, also presented questions in blocks (“all” questions appeared in one block; “some” questions appeared in another, with order of blocks counterbalanced). In building on this work, we attempted to keep aspects of the design comparable.

Procedure. Children were tested individually in a quiet room at their schools. Before the testing session, each child was told that he or she would be playing a special test with other children. Most of the 3-year-olds (14 of 18) then received a warm-up task to introduce them generally to answering questions posed by the experimenter. The warm-up consisted of four simple yes/no questions (e.g., “Is this a spoon?” when showing a picture of a spoon; “Is this a chair?” when showing a picture of a pig); two of the correct answers were “yes” and two of the correct answers were “no.” Children performed extremely well on the warm-up task.

Children were then introduced to the main task. The experimenter said, “We are going to play a game with this special board [an 8 × 10 in. photo frame covered in Velcro] and these special cards [colored, laminated 3 × 5 in. index cards with Velcro on the back] that you can stick right on the board. Each card has a question on it, and you get to pick a card and stick it right on the board, and then we can talk about what it says on the card. Okay? The cards were color-coded by trial block—yellow for generic green for “all,” and purple for “some.” Cards for each block were kept in separate envelopes. The experimenter emptied an envelope of cards for a given block, laid them out face down before the child, and told the child that he or she could pick one card at a time and put it up on the board. For each trial, the child picked a card and placed it in the photo frame. The experimenter read the question aloud and both audiotaped and wrote down the child’s response.

Three-year-olds were also given a posttest task designed to test their comprehension of “all” and “some” with concrete and immediately perceptible referents (modeled on a pretest developed by Smith, 1979, p. 439). Children were told, “Now we have a new game to play for a little while. Look, I have a box and crayons.” The experimenter produced a small clear container and a plastic bag containing four crayons. There were eight trials, four with “all” and four with “some.” For each trial, the experimenter placed zero, two, three, or four of the crayons into the box and then asked, “Are all of the crayons in the box?” (on “all”) trials or “Are some of the crayons in the box?” (on “some”) trials. The questions were presented in two blocks, an “all” block and a “some” block, with order of the blocks counterbalanced across participants. The order of the questions within a block (zero, two, three, or four crayons in the box) was randomized separately for each participant, with the constraint that the potentially ambiguous questions (zero for “all” and four for “some” were always presented last in the block). Smith (1979) had found that children as young as 4 years of age had no difficulty with this task, and so we included only 3-year-olds on the posttest.

Adults were tested in a group with a paper-and-pencil version of the task. They were told that they would be participating in an experiment that was designed for young children. They received a booklet with one question on each page. They were asked to write down their answer to each question in the space provided below the question itself.

Results

Responses were scored as follows: 1 for each “yes” response, 0 for each response of “no” or “none,” and 0.5 for each response that didn’t fit into either category (when respondents did not give a yes or no answer; e.g., when asked “Do girls have curly hair?” a child might reply “some do” without responding “yes” or “no”). Such cases were relatively rare, accounting for only 4.6% of the responses of 3-year-olds, 6.6% of the responses of 4-year-olds, and 1.3% of the responses of adults.2 The primary reason for coding these responses as 0.5 is that they indicate a reluctance to commit to a definitive “yes” or “no” answer. A 0.5 coding reflects this, as it indicates an answer that is intermediate between “yes” and “no.” A further motivation is that this coding scheme is unbiased, weighted toward neither response.

For each participant, we summed the scores, separately for each of the nine cells in the study design (three forms of the question—all—generic—some) × three types of properties (wide scope, narrow scope, irrelevant)). Thus, each participant received nine scores, each of which could range from 0 to 4 (for 4-year-olds and adults) or from 0 to 2 (for 3-year-olds). The scores of the 3-year-olds were then transformed by multiplying each score by two, thereby rendering them on the same scale as the two older age groups.

These scores were then entered into a 3 (age: 3 years, 4 years, or adult) × 3 (wording: all, generic, or some) × 3 (property: wide scope, narrow scope, irrelevant) analysis of variance (ANOVA). The results can be seen in Figure 1. As expected, there was a significant main effect of property, F(2, 142) = 845.14, p < .0001. Not surprisingly, participants endorsed wide-scope properties most often (M = 10.63 out of 12), narrow-scope properties less often (M = 6.81 out of 12), and irrelevant properties least often (M = 0.53 out of 12).

All the remaining effects involved wording. First, there was a significant effect of wording, F(2, 142) = 46.80, p < .0001, with higher scores for the “all” wordings than for the “generic” wordings. Second, there was a significant interaction of wording and age, F(4, 142) = 46.00, p < .0001. Separate ANOVAs in each age group revealed the following: for the 3-year-olds, a significant main effect of wording, F(2, 26) = 142.80, p < .0001; for the 4-year-olds, a significant main effect of wording, F(2, 70) = 101.34, p < .0001; and for adults, a significant main effect of wording, F(2, 46) = 204.20, p < .0001. For all three age groups, scores were highest on the “all” wordings, followed by the “some” wordings, and lowest on the “generic” wordings.

1 Three of the 4-year-old children received the task without the Velcro frame, but all other aspects of the procedure were the same as with the other child participants.

2 Approximately three fourths of these intermediate responses occurred in three of the nine cells of the study design: “all” questions regarding narrow-scope properties (e.g., “Do all clothes have zippers?”), generic questions regarding narrow-scope properties (e.g., “Does milk have chocolate in it?”), and “some” questions regarding wide-scope properties (e.g., “Are some fires hot?”), accounting for 22%, 30%, and 24% of the intermediate responses, respectively. These cells can be considered anomalous in that they pair either a noun phrase typically broad in scope with a narrow-scope property or a noun phrase typically narrow in scope with a broad-scope property. The rest of the indeterminate answers were scattered fairly equally across the other six cells of the study design, each cell accounting for 0% to 6% of these responses.
separately. Doing so revealed a striking developmental change from 3 to 4 years of age.

At age 3, children did display a significant effect of property, $F(2, 34) = 113.84, p < .0001$. However, wording exerted no significant effects. In contrast, for 4-year-olds, there were significant effects of wording, $F(2, 34) = 9.66, p < .001$; property, $F(2, 34) = 217.94, p < .001$; and Wording $\times$ Property, $F(4, 68) = 20.53, p < .001$. With the wide-scope properties, 4-year-olds were more likely to answer “yes” in response to “all” questions, than in response to “some” questions ($p < .05$). There was no significant difference between generic and either “all” or “some” on these items. In contrast, with narrow-scope properties, 4-year-olds were more likely to answer “yes” in response to “some” and generic questions than in response to “all” questions ($p < .01$). There was no significant difference between “some” and generic on these items. Finally, for both generic and “all” questions considered separately, 4-year-olds were more likely to affirm wide-scope properties than narrow-scope properties ($p < .05$). In contrast, there was no significant difference between wide-scope and narrow-scope properties for “some” questions.

Adults showed patterns very similar to those of 4-year-olds. There were significant effects of wording, $F(2, 74) = 177.61, p < .0001$; property, $F(2, 74) = 942.35, p < .0001$; and Wording $\times$ Property, $F(4, 148) = 119.41, p < .0001$. With the wide-scope properties, adults were more likely to answer “yes” in response to “all” questions than in response to generic questions ($p < .01$). There was no significant difference between “all” and generic or between “all” and “some” on these items. In contrast, with narrow-scope properties, adults were more likely to answer “yes” in response to “some” and generic questions than in response to “all” questions ($p < .001$) and were more likely to answer “yes” in response to “some” questions than in response to generic questions ($p < .001$). Finally, for both generic and “all” questions considered separately, adults were more likely to affirm wide-scope properties than narrow-scope properties ($p < .001$). In contrast, for “some” questions, adults showed the opposite pattern: They were more likely to affirm narrow-scope than wide-scope properties ($p < .05$).

When we compared the age groups with one another, using Tukey’s honestly significant difference test, we found that all of the age differences were localized in two types of questions: narrow-scope properties in the “all” condition (with 3-year-olds’ responses higher than 4-year-olds’, $p < .001$, which were higher than adults’, $p < .005$), and narrow-scope properties in the “some” condition (with adults’ responses higher than those of either 3-year-olds or 4-year-olds, $p < .005$, which did not differ from one another). There were no significant age effects in the generic wording condition.

Order effects. Because the order of wording was blocked (i.e., all generic questions in one block, all “all” questions in another block, all “some” questions in another block, with blocks quasi-counterbalanced), this raises the possibility that children may have developed response sets within each block that led to artificially distinguishing the three kinds of wording. Note that the task could not have created a distinction where none exists. That is, if children and adults come into the study with no a priori differentiation between generic, “all,” and “some,” then hearing such items blocked and contrasted would not impose a novel semantic interpretation on the spot. In particular, it is implausible that partici-

Figure 1. Mean number of “yes” responses (out of 4 possible) in Study 1. Top: results for 3-year-olds ($N = 18$). These scores were doubled to be on the same scale as those of the 4-year-olds and adults. Middle: results for 4-year-olds ($N = 18$). Bottom: results for adults ($N = 38$).
pants would reach a consistent interpretative distinction between the three forms simply on the basis of noticing that the blocks differed. Nonetheless, the blocking of trials may have influenced response patterns, and it is important to examine whether the results hold up when possible carryover effects are eliminated.

To examine this issue, we conducted a secondary analysis that reran Study 1 as a between-subjects design, by including only the data from the first block of trials. Figure 2 presents the results of this analysis. Keep in mind that this is a highly stringent test, as for the 3- and 4-year-olds there were only 6 participants per wording condition on this analysis. Specifically, we conducted a 3 (age: 3-year-olds, 4-year-olds, or adults) × 3 (wording condition: all, generic, or some) × 3 property scope (wide scope, narrow scope, or irrelevant) ANOVA, with age and wording condition as between-subjects variables and property scope as a repeated-measures variable. From this analysis, we again obtained the expected main effect of property, F(2, 130) = 329.91, p < .001, and a Property × Age interaction, F(4, 130) = 2.69, p < .05. All remaining effects involved wording. There was a main effect of wording, F(2, 65) = 11.79, p < .001; a Wording × Age interaction, F(4, 65) = 3.37, p < .02; and a three-way interaction involving wording, age, and property type, F(8, 130) = 4.78, p < .001.

To explore this three-way interaction in detail, we conducted separate Wording × Property ANOVAs within each age group separately. As in the overall ANOVA reported earlier, 3-year-old children displayed a significant effect of property, F(2, 30) = 56.00, p < .001, but no significant effects involving wording. In contrast, for 4-year-olds, there were significant effects of wording, F(2, 15) = 3.47, p = .06; property, F(2, 30) = 65.58, p < .001; and Wording × Property, F(4, 30) = 7.81, p < .001. With the narrow-scope properties, 4-year-olds were most likely to answer “yes” in response to “some” questions, least likely to answer “yes” in response to “all” questions, and responses to generic questions were intermediate (p < .05). Finally, for both generic and “all” questions considered separately, 4-year-olds were more likely to affirm wide-scope properties than narrow-scope properties (p < .05). In contrast, there was no significant difference between wide-scope and narrow-scope properties for “some” questions.

Adults showed patterns very similar to those of 4-year-olds. There were significant effects of wording, F(2, 35) = 80.95, p < .001; property, F(2, 70) = 340.01, p < .001; and Wording × Property, F(4, 70) = 46.74, p < .001. With narrow-scope properties, adults were more likely to answer “yes” in response to “some” and generic questions than in response to “all” questions (p < .001) and were more likely to answer “yes” in response to “some” questions than in response to generic questions (p < .001). Finally, for both generic and “all” questions considered separately, adults were more likely to affirm wide-scope properties than narrow-scope properties (p < .02). In contrast, for “some” questions, there was no significant difference between wide-scope and narrow-scope properties.

Thus, the results of this reanalysis largely replicate those of the overall study, with 3-year-olds demonstrating only a main effect of property type and no effects of wording but 4-year-olds and adults each showing a main effect of property type, a main effect of wording, and a Property Type × Wording interaction. Furthermore, just as in the overall analysis, this analysis demonstrated that 4-year-olds and adults (but not 3-year-olds) treat generics as partly like “all” and partly like “some.” On the narrow-scope properties, 4-year-olds and adults say “yes” to generic questions significantly more often than to “all” questions but significantly less often than to “some” questions. In other words, responses to questions like “Do girls have curly hair?” are intermediate between “Do all girls have curly hair?” and “Do some girls have curly hair?” Altogether, these subsidiary analyses render it implausible that blocking introduced artificial response sets.

**Figure 2.** Reanalysis of Study 1 results, including only the first block of trials. The dependent measure is the mean number of “yes” responses (out of 4 possible). Top: results for 3-year-olds (N = 18). These scores were doubled to be on the same scale as those of the 4-year-olds and adults. Middle: results for 4-year-olds (N = 18). Bottom: results for adults (N = 38).
“Yes, some girls do”) or with “all” (e.g., in response to “Are fires hot?” a child might say “All fires are” or “Yes, all fires are”). We did not include as qualifications those cases in which a participant simply repeated information in the question (e.g., “Are all fires hot?” “Yes, they’re all hot.”). Thus, qualifications, by definition, supplied information that could not be assumed in the question.

As expected, most “some” qualifications were given in response to narrow-scope properties (90%), and most “all” qualifications were given in response to wide-scope properties (96%). We therefore included only those qualifications that were provided for the expected scope (i.e., “some” for narrow-scope properties and “all” for wide-scope properties). However, it is important to note that the patterns of results reported below are largely equivalent when all qualifications are analyzed.

Our main question is what sorts of qualifications children and adults provided for generics and how these compared to their qualifications for the other wording conditions. Because the different wording conditions differed in terms of which qualifications were possible (see above), we conducted a series of paired t tests comparing qualifications of generic responses with qualifications of either “all” or “some” responses, within each age group separately (see Table 1). As can be seen, 3-year-olds did not distinguish between generics and either “all” or “some” in the qualifications they provided. Both 4-year-olds and adults, however, showed distinct patterns of response: more “some” qualifications for generic questions than for “all” questions, and more “all” qualifications for “some” questions than for generic questions. The fact that generic questions are more often qualified with “some” indicates that generics do not themselves imply “some.” Overall, this pattern suggests that generics are distinct from both “all” and “some.”

Individual response patterns. The finding that generics were in between “some” and “all” raises the question of whether individual participants consistently treated generics as intermediate or whether instead this finding is the result of averaging across participants. To address the issue, we examined the response patterns of individual participants. Because the data were complex, and because the wording effects were most pronounced on the narrow-scope properties, we focused specifically on this subset of the data. (Wide-scope properties tended to yield ceiling effects, and irrelevant properties yielded floor effects.) We coded each participant’s responses to the narrow-scope properties into one of four categories: (a) Generic = intermediate. These participants had higher scores for “some” than for generic and higher scores for generic than for “all” (“some” > generic > “all”). In other words, they treated generic as intermediate between “some” and “all.” (b) Generic = “all.” These participants had higher scores for “some” than for generic and “all” and had equivalent scores for generic and “all” (“some” > generic, “all”). In other words, they differentiated “some” from “all” and treated generic as equivalent to “all.” (c) Generic = “some.” These participants had higher scores for “some” and generic than for “all” and had equivalent scores for “some” and generic (“some,” generic > “all”). In other words, they differentiated “some” from “all,” and treated generic as equivalent to “some.” (d) Other. This included all other response patterns.

Results are shown in Table 2. For both 4-year-olds and adults, the most frequent response pattern was to treat generics as intermediate between “all” and “some.” Among those who did not, some participants treated generics as equivalent to “some,” and some treated generics as equivalent to “all,” though the former was slightly more common. Most of the 3-year-olds showed “other” response patterns, primarily because most of them did not differentiate “all” and “some.”

Item analyses. We also examined the results for items considered individually to determine the generalizability of the results across items. As we did in the individual response analysis, we focused solely on the narrow-scope property items. There were 12 narrow-scope items for 4-year-olds and adults and 6 narrow-scope items for 3-year-olds. We coded responses to each of these items into the same categories as those described above for the individual response patterns: (a) generic = intermediate (“some” > generic > “all”), (b) generic = “all” (“some” > generic, “all”), (c) generic = “some” (“some,” generic > “all”), and (d) “other.”

Three-year-olds showed little consistency: 1 item showed Pattern a, 2 showed Pattern b, 1 showed Pattern c, and 2 showed no pattern (i.e., Pattern d). For 4-year-olds, however, the items showed much greater consistency: Fully half the items (6 out of 12) showed Pattern a, whereby generics are intermediate in scope between “all” and “some.” Three of the items showed generics as equivalent to “some,” and the remaining 3 items showed no consistent pattern. Among the adults, all 12 items showed Pattern a, in which generics were intermediate in scope between “all” and “some.”

Posttest. The posttest was given to 3-year-olds only, to test their understanding of “all” and “some” in a concrete context with immediately perceptible referents. Because of experimenter error, 2 of the 3-year-olds did not receive the posttest, so a total of 16 children were included in the analyses. For each participant, we scored a response of “yes” as 1 and a response of “no” as 0 (see Table 3 for results). Overall, children performed extremely well on this task. Children typically applied “all” just to the situation in which four of the four crayons were in the box. In contrast, they typically applied “some” to the situations in which two, three, or (less often) four crayons were in the box.

Table 1
Mean Number of Qualifications Using “Some” or “All” (Out of 12 Possible) in Study 1

<table>
<thead>
<tr>
<th>Question</th>
<th>3-year-olds</th>
<th>4-year-olds</th>
<th>Adults</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Some” qualifications</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>“All”</td>
<td>1.11</td>
<td>0.83</td>
<td>0.11</td>
</tr>
<tr>
<td>Generic</td>
<td>0.78</td>
<td>1.56</td>
<td>1.71</td>
</tr>
<tr>
<td>“All” qualifications</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>“Some”</td>
<td>0.00</td>
<td>1.61</td>
<td>0.37</td>
</tr>
<tr>
<td>Generic</td>
<td>0.00</td>
<td>0.33</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Note. Scores for 3-year-olds were doubled to be on the same scale as those of the 4-year-olds and adults. Subscript a indicates significant difference between “all” and generic questions, p < .05; one-tailed; subscript b indicates significant difference between “all” and generic questions, p < .01; subscript c indicates significant difference between “some” and generic questions, p < .10; subscript d indicates significant difference between “some” and generic questions, p < .05.
We entered the scores into a 2 (wording: all or some) × 4 (number: zero, two, three, or four) ANOVA. Results indicated clear effects of wording, $F(1, 15) = 45.00, p < .001$; number, $F(3, 45) = 30.80, p < .001$; and Wording × Number, $F(3, 45) = 24.45, p < .001$. When comparing “all” and “some” trials directly with paired $t$ tests, we found that children significantly distinguished the two kinds of questions on trials asking about two out of four crayons ($p < .001$), three out of four crayons ($p < .01$), and four out of four crayons ($p < .05$). Furthermore, individual children’s response patterns also suggested that these children had consistent understanding. Eleven out of the 16 children showed the correct pattern on “all” (i.e., “yes” to four out of four only), and 12 out of the 16 children showed the correct pattern on “some” (i.e., “no” to zero out of four and “yes” to two out of four and three out of four; for the purposes of this analysis, we did not include responses to the four out of four trials, given the ambiguity of “some” on these questions).

**Discussion**

Overall, these results are consistent with a semantic analysis in which by which children are 4 years of age, generics imply broad generalizations but also allow for exceptions. In other words, 4-year-olds and adults interpret generics as being reducible to neither “all” nor “some.” Like “all,” generics are more appropriate for wide-scope generalizations (e.g., “Fires are hot”) than for narrow-scope properties (e.g., “Girls have curly hair”). Yet like “some,” generics are nonetheless endorsed above chance for narrow-scope properties (e.g., “Girls have curly hair”). Individual response patterns further reveal that most 4-year-olds and adults consistently treat generics as intermediate between “all” and “some” in meaning. However, in their spontaneous qualifications, 4-year-olds and adults treat generics as more like the quantifier “all.”

These distinctions (between generic and “all” and between generic and “some”) are subtle and so provide a powerful test of children’s grasp of generic implications. First, the distinction between “all” and generics emerges despite the fact that both noun phrase types imply the category as a whole, differing only in their commitment to every instance. At times generics are indistinguishable from “all,” being used to convey properties true of every category member (e.g., “Dogs are mammals”). Nonetheless, generics allow for the possibility of counterexamples in a way that “all” does not (e.g., “Dogs have four legs” is not invalidated by the existence of three-legged dogs). Second, the distinction between generics and “some” is subtle in that the bare plural form has dual functions, at times referring to generics (e.g., “Bears like to eat ants”) and at times referring to an indefinite plural (e.g., “I saw bears in the park yesterday”). Note that the indefinite plural is comparable to a noun phrase with “some” (i.e., “I saw bears in the park yesterday”) is comparable in meaning to “I saw some bears in the park yesterday”). Given these potential confusions—on the one hand between generics and “all,” and on the other hand between generics and “some”—it is all the more impressive that 4-year-olds distinguish among the three forms.

We found that, in striking contrast to the data from 4-year-olds and adults, 3-year-olds did not distinguish “all,” generic, and “some.” Their difficulty cannot be attributed to lack of knowledge (e.g., not knowing that fires are hot or that girls can have either curly hair or straight hair), because they did consistently distinguish wide from narrow from irrelevant properties. For example, they were more likely to endorse the proposition that fires are hot than that girls have curly hair. Instead, the problem is specifically with differentiating among the three types of wording. For example, they were statistically as likely to accept that all girls have curly hair as they were to accept that some girls have curly hair.

However, it is also not the case that 3-year-olds know nothing about the meanings of “all” and “some,” because they performed extremely well on the all–some posttest, which used a small set of concrete, available objects. Instead, 3-year-olds had difficulty extending this knowledge beyond the realm of instances that were immediately present to a consideration of categories in the abstract.

**Study 2: Elicited Production Task**

Study 2 was similar in logic to Study 1 but differed in design. Here, children were asked to produce their own utterances, under the guise of giving information to Zorg, an alien puppet from outer space. The prompts children heard used one of three kinds of cues: generic (e.g., “What can you tell Zorg about dogs?”), “all” (e.g., “What can you tell Zorg about all dogs?”), or “some” (e.g., “What can you tell Zorg about some dogs?”). An analogous paper-and-pencil version of the task was developed for adults. Of interest was whether the responses that children and adults generated in the three types of wording conditions would differ in scope and content. We predicted that properties generated under “all” prompts would be broadest in scope, properties generated under

**Table 2**

<table>
<thead>
<tr>
<th>Response pattern</th>
<th>3-year-olds</th>
<th>4-year-olds</th>
<th>Adults</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generic = intermediate</td>
<td>6 (1)</td>
<td>56 (10)</td>
<td>68 (26)</td>
</tr>
<tr>
<td>Generic = some</td>
<td>17 (3)</td>
<td>17 (3)</td>
<td>24 (9)</td>
</tr>
<tr>
<td>Generic = all</td>
<td>11 (2)</td>
<td>6 (1)</td>
<td>8 (3)</td>
</tr>
<tr>
<td>Other</td>
<td>67 (12)</td>
<td>22 (4)</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>

Note. Scores indicate the percentage of participants in each age group adhering to one of three response patterns or falling within no consistent response pattern (Other). The numbers of participants are provided in parentheses.

**Table 3**

<table>
<thead>
<tr>
<th>Score</th>
<th>“All”</th>
<th>“Some”</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 out of 4</td>
<td>.06†</td>
<td>.13†</td>
</tr>
<tr>
<td>2 out of 4</td>
<td>.13†</td>
<td>.94*</td>
</tr>
<tr>
<td>3 out of 4</td>
<td>.06†</td>
<td>.94*</td>
</tr>
<tr>
<td>4 out of 4</td>
<td>.94*</td>
<td>.69</td>
</tr>
</tbody>
</table>

Note. Maximum possible score was 1. † Below chance by one-sample $t$ test ($p < .005$). * Above chance by one-sample $t$ test ($p < .001$).
“some” prompts would be narrowest in scope, and properties generated under generic prompts would be intermediate in scope. This is a more demanding task than the yes–no task of Study 1. Instead of simply assessing the truth value of a series of utterances, children were required to generate properties on the basis of linguistic cues alone. Good performance on this task requires that children have sufficient access to generic knowledge to produce it on demand. It also requires that they demonstrate sensitivity to the linguistic prompts and resist any tendency they may have to answer all questions in terms of either stereotypes (e.g., prototypical dogs) or salient exemplars (e.g., their own pet dog). Altogether, it provides a stronger test of children’s attentiveness to generic language. Given the demands of this task, and given 3-year-olds’ lack of condition effects in Study 1, the youngest age group we tested was on average between 4 and 5 years of age.

Method

Participants. Forty-eight children (age range = 4 years 1 month to 6 years; mean age = 4 years 10 months—henceforth referred to as 4-year-olds) and 37 adults participated in the main experiment. In addition, 92 adults were involved in ratings of generated properties. The adults were undergraduates at a large midwestern university; the children attended local preschools in the same town as the university.

Items. There were 12 categories in this experiment: birds, cats, dogs, fish, boys, girls, smart people, shy people, tables, shirts, cars, and either computers (for adults) or houses (for children). The items included a variety of natural kinds, social categories, and artifacts in order to provide a broad test. We had originally gathered the data from adults, including the category “computers,” but pilot testing with children revealed that they had relatively little knowledge about computers. Therefore, we replaced this item with “houses” (also a complex artifact) for the younger age group.

Procedure. Each participant was randomly assigned to one of the three conditions: “all,” generic, or “some,” which varied between participants. The details of the procedure differed for adults versus children, so they are described separately below.

The procedure for adults was as follows. Each participant was tested in a small group. Participants received a page of written instructions, using the category “trees” (or “all trees” or “some trees,” depending on the wording condition) as an example. The instructions simply asked participants to list five things that were true of each category. Then they received a booklet containing 12 pages, with 1 page per category. The pages were randomized, with the constraint that every three items include one animal category, one artifact category, and one social category, in random order. As a consequence, no more than two items in a row were from the same domain. For each page, participants were asked to list five properties of the category (e.g., “List five things that are true of dogs” [generic], “List five things that are true of some dogs” [some], or “List five things that are true of all dogs” [all]). Participants worked through the booklet at their own pace.

The procedure for children was modified from that for adults, as pilot testing revealed that children were reluctant to say much when simply asked to list properties for the experimenter. We found that children were much more forthcoming when asked to give information to Zorg, an alien puppet from outer space, who was introduced as not knowing very much about things here on Earth. The prompts children heard used one of three kinds of cues: generic (e.g., “What can you tell Zorg about dogs?”), “all” (e.g., “What can you tell Zorg about all dogs?”), or “some” (e.g., “What can you tell Zorg about some dogs?”). After the initial prompt, we also asked a set of three additional prompts for each item to elicit as much information as possible. These questions concerned (a) what Xs do (for animals and people) or what you do with Xs (for artifacts), (b) what Xs have, or (c) what Xs wear (for people) or what color Xs are (for nonpeople). In each question X was replaced with the appropriate phrase (e.g., dogs, all dogs, or some dogs, depending on the wording condition.) To keep children’s interest and attention, after they were interviewed about a given item, they were shown a Velcro-backed line drawing of that item, which they were allowed to place on a Velcro picture frame. Items were presented in random order, with the constraint that no more than two items in a row be from the same domain (animals, artifacts, or social categories). The entire procedure with children was audi-taped.

Results

Coding. Transcripts of the elicitation sessions were prepared, and coding proceeded in a series of steps. First, each utterance was broken down into propositions (e.g., “They play with string and chase a mouse” was considered two separate propositions: play with string or chase a mouse). Next, for the children’s transcripts, each proposition was coded for consistency between the linguistic form of the experimenter’s prompts and the linguistic form of the child’s response. We coded consistency as a means of excluding trials on which children clearly ignored or contravened the instructions. For example, if a participant received a generic prompt about “tables” but gave a property that was prefixed with “Some tables . . .,” this was considered inconsistent. It was not surprising that children sometimes countered the experimenter’s question with information differing in form, as this is part of the normal give-and-take of conversation. However, for the purposes of experimental investigation, it is important to exclude trials on which children were (in effect) answering a question other than the one the experimenter asked. We were conservative in excluding trials and did so only when the child explicitly marked that his or her response differed in form from that of the experimenter’s prompt. In many cases the child simply provided a property without any qualification (e.g., Experimenter: “Could you tell Zorg what color all dogs are?” Child: “Brown.”); such trials were assumed to be consistent. (Note that this assumption may overestimate consistency, but doing so would only work against our hypotheses.) On this measure, 80% of the propositions were consistent with the experimenter-provided prompts (see Figure 3). There was a tendency for children to produce fewer consistent properties in the “all” condition, which may reflect the greater difficulty of generating properties that lack exceptions.

One main coder broke down the transcripts into propositions and coded for consistency for all the children (N = 48). A second coder coded 25% of the children’s data (12 of the 48 participants). Agreement on identifying propositions was 95.5%; agreement on
consistency coding was 97%. Disagreements were resolved by discussion.

Scope ratings. All the remaining analyses focus just on those utterances that were consistent. It is important to reiterate that consistency is defined in terms of linguistic form, not linguistic content. For example, a child who is asked “What color are all dogs?” and responds “brown” would be coded as providing a response consistent with the “all” prompt even though the content of the property is true of less than all members. Our empirical question was whether children generated properties differing in scope depending on the linguistic prompt, focusing just on those trials in which they complied with the experimenter’s instructions (which, as noted above, was the vast majority of trials).

Our next step was to ask an independent group of adults to rate the frequency with which each consistent property was true of the category in question. Specifically, they were asked to indicate how often (using a scale ranging from 0%–100%, in 10% increments) each property in the list was true for the category as a whole. Adult raters were blind to the purpose of the ratings, to the linguistic prompts that were provided by the experimenter, and to the linguistic format in which the property was expressed. (So, e.g., “some dogs are furry,” “all dogs are furry,” and “dogs are furry” would all be rendered as “are furry” for the adult raters.)

Because of the large number of properties that were generated, each adult rated properties from only 3 of the 12 categories (one animal kind, one social kind, and one artifact category): specifically, the full set of either child-generated or adult-generated properties for 3 categories. Altogether, there were 38 raters for adult properties ($N = 9–10$ raters per category) and 54 raters for child properties ($N = 13–14$ raters per category).

After obtaining the full set of adult ratings, we then averaged the adult ratings for each property. For example, the ratings for the property “bark at the parking lot,” which had been generated by one of our child participants for the category “dogs,” were averaged across the 14 adults who provided ratings for that property. We then went back and assigned this mean score for the trial on which the child participant provided this particular property. We then determined for each participant in the main study the average of these mean ratings assigned to the properties they generated, for each of 12 categories.

This rather complex process can be better understood by considering an example from one of our child participants. This child (J.) was in the generic condition and generated five properties about birds: have wings, fly, eat worms, have a beak, and have two feet. Each of these properties was rated by 13 adult raters, who generated ratings that averaged to the following: 93.46 (have wings), 87.31 (fly), 61.54 (eat worms), 91.54 (have a beak), and 82.31 (have two feet). When the ratings of these five properties were averaged, we obtained an overall score of 83.23 for “birds” for Child J. This process was then repeated for each of the 12 categories for each participant so that each child and each adult ended up with 12 scores, one per category.

Primary analysis. For the primary analysis, we averaged over the 12 categories, which resulted in a single score for each participant. Our primary analysis consisted of a $2 \times 3$ (age: preschool or adult) ANOVA, with the single averaged score as the dependent measure. Both age and condition were between-subjects variables.

Results are shown in Figure 4. There was a significant main effect of wording, $F(2, 79) = 33.94, p < .001$; a main effect of age group, $F(1, 79) = 92.76, p < .001$; and a Wording × Age Group interaction, $F(2, 79) = 3.53, p < .05$. Results indicated that both children and adults were sensitive to the linguistic prompts provided by the experimenter. In both age groups, “some” elicited properties of lesser scope than “all” or generic, which did not differ from one another. However, adults overall provided properties that were broader in scope than those provided by the children. The significant two-way interaction indicates that the age difference was greatest in the “all” condition.

Item analyses. Particularly given the breadth of items (encompassing natural kinds, social categories, and artifacts), we wished to examine whether the effects held up over the items taken as a whole and were not restricted to a small subset. For each of the 12 items, we averaged responses across all participants of a given age group and wording condition and then compared responses across wording conditions. For children, “all” was greater than “some” on 10 of the 12 items, generics were greater than “some” on 12 of the 12 items, but surprisingly “all” was less than generics on 11 of the 12 items. In contrast, for adults, all 12 items showed the pattern: “all” > generic > “some.” Thus, for both children and adults, there was considerable consistency over items. “All” was consistently treated as wider in scope than “some,” and generics were consistently treated as wider in scope than “some.” However, there was also a developmental shift in the relation between “all” and generics, with children tending to treat generics as wider in scope and adults treating “all” as wider in scope.

Property content. In addition to examining property scope, we were interested in examining property content. Our question was whether there are qualitative differences between “all,” generic, and “some”—not just quantitative differences. To address this question, we first coded each of the properties generated into one of five coding categories: physical (physical part, appearance, or bodily feature; e.g., “have a nose”), action (physical behavior, or in the case of inanimate objects, what something does [in active sense]; e.g., “eat worms”), function (use for someone or something else; e.g., “write on them”), mental state (psychological states including thoughts, beliefs, and emotions; personality feature; such as “like tools”), and other (all else). These coding categories were determined on the basis of intuitive partitioning of the responses produced. Within each age group, coding was conducted by two coders who were blind to the wording condition in which the property was generated. (For adult-generated properties, one coder coded all properties, and the second coded half the properties; for child-generated properties, the two coders each coded all the properties.) Agreement was 96% for adult-generated properties and 85% for child-generated properties. Disagreements were resolved by discussion.

For each of the 12 categories, each participant received 4 scores, representing the total number of physical, action, function, and

---

3 This step was not conducted for the adults’ transcripts, for two primary reasons: Adults typically simply listed properties without providing syntactic subjects, thus making it impossible to check what form adults had in mind, and adults were assumed to be more compliant with the experimenter’s instructions, given the formal nature of the paper-and-pencil testing session and adults’ vaster testing experience.
mental codes, divided by the total number of properties generated for that category. (Percentage scores were required rather than mean scores, because children generated a variable number of properties.) Then, for each of the 4 coding categories considered separately, we averaged the 12 scores (from the 12 categories tested) to obtain a single score for each participant. All told, each participant thus received four composite scores: physical (averaged over the 12 items), action (averaged over the 12 items), function (averaged over the 12 items), and mental (averaged over the 12 items). The physical and action scores accounted for the majority of responses for both children and adults (physical: 43% of children’s responses, 41% of adults’ responses; action: 33% of children’s responses, 23% of adults’ responses). The remaining responses appeared consistently less often (ranging from 4%–16% of responses at each age). Therefore, analyses of property content focused exclusively on properties with physical or action content.

We conducted two ANOVAs, one for each composite score, with age (4-year-olds or adults) and wording condition (all, generic, or some) as between-subjects factors. For the physical codes, the ANOVA revealed a main effect of condition, $F(2, 79) = 7.95$, $p < .01$, and no main effect or interaction due to age. The generic condition resulted in a lower proportion of physical properties ($M = 36\%$) than either the “all” condition ($48\%$) or the “some” condition ($44\%$). For the action codes, the ANOVA revealed main effects of condition, $F(2, 79) = 3.37$, $p < .05$, and of age, $F(1, 79) = 23.73$, $p < .001$. Children generated a higher proportion of action properties than adults (33% and 23%, respectively). There was also a tendency for action codes to be higher in the generic condition (32%) than in either the “all” condition (27%) or the “some” condition (28%).

Overall, the content analyses suggest that generics tend to be differentiated from both “all” and “some,” for both children and adults. For 4-year-olds, generics elicited fewer physical properties and, to a lesser extent, more action properties; for adults, generics elicited fewer physical properties and more action properties. These results suggest that properties elicited in generic form differ in content, not just scope, from those elicited with either the quantifier “all” or the quantifier “some.”

Discussion

On the elicited production task, as on the comprehension task in Study 1, preschool children and adults distinguished generics from “some” in scope. Generics are consistently broader in scope than “some” statements. Furthermore, adults but not children treat generics as narrower in scope than “all” statements. It is particularly striking that children displayed sensitivity on this more demanding task, which required that they come up with properties on their own.

In contrast to the analyses focused on property scope, analyses focused on property content reveal that generics consistently differ from both “all” and “some,” which did not differ from one another. This result provides further support for the claim that generics are distinct from “all” and “some” statements. Surprisingly, however, generics are not intermediate on this analysis but rather display a unique pattern. Prasada (2000) has suggested that generics reflect essential properties, and we speculate that perhaps essential properties are somewhat less likely to be outward perceptual features and more likely to be actions and mental states.

General Discussion

The results of these studies are clear: 4-year-old children and adults distinguish between generics and statements using “all” or “some.” We note that in both cases (generic vs. “all” and generic vs. “some”) the distinction is a subtle one. Generics are often characterized as being generally true, yet even so children honor a distinction between generics and the more inclusive “all.” On the other hand, the distinction between generics and “some” is upheld despite the fact that the same form of the noun phrase (bare plural) is used in both generic utterances (e.g., “Bears are ferocious”) and indefinite utterances (e.g., “I saw bears at the zoo yesterday”). Thus, the semantic features of generics presented in the introduction to this article (neither reducible to “all” nor reducible to indefinite “some”) are appreciated by adults and preschool children alike.

One important interpretive issue from Study 1 needs to be addressed. Namely, it is possible that the intermediate pattern of responses to Study 1 (with generics receiving a number of “yes” responses, that is, in between their responses to “all” and their responses to “some”) could indicate confusion or mixed responses rather than a consistent generic interpretation. However, several additional pieces of information argue against this alternative account. First, the 4-year-olds in Study 1 show a response pattern that is in all respects quite similar to that of adults on the same task. Second, children’s spontaneous qualifications likewise are consistent with the construal that children distinguish generics from “all” and “some.” Third, the data from Study 2 further bolster the idea that children treat generics as distinct from “some,” this time using a production task rather than a yes–no comprehension task. Finally, the findings from other published reports using different research methods (Gelman & Raman, in press; Gelman et al., 2002) are consistent with the interpretation provided for the results of Study 1.

It is somewhat surprising that on the semantic interpretation (yes–no) task of Study 1, children treated generics as intermediate between “all” and “some,” but on the elicited production task of Study 2, children treated generics as equivalent to “all.” However, we suspect that this may largely reflect the difficulty of producing properties that are true of all members of a category (e.g., Mervis & Rosch, 1981). One would be hard pressed to list features that are true of all shirts, for example, other than that they are worn (which is typically listed in the generic condition as well). Furthermore,
the content analyses demonstrate that generics are distinct from "all" as well as "some" even in Study 2. Overall, then, both studies confirm that generics are clearly distinguished from both a "some" interpretation and an "all" interpretation.

When does this understanding emerge? One important finding is that generics are understood early in development, by 4 years of age. This is a surprising finding given the subtlety of the semantic distinctions under consideration. However, there is a second finding as well, namely, that performance changes rather abruptly between 3 and 4 years of age. In Study 1, 3-year-olds displayed no sensitivity to the distinction between "all," generics, and "some" on the main comprehension task, whereas 4-year-olds were notably adult-like in their responses.

Despite children’s failure to differentiate among the linguistic cues, these results do not demonstrate that 3-year-olds lack generic understanding. Indeed, in Study 1 the overall means for generics do not change significantly with age (see Figure 1). Three-year-olds’ response patterns on the generic questions are statistically indistinguishable from those of 4-year-olds and adults. All of the significant age changes are with "all" and "some." We therefore propose the following interpretation: Under certain conditions (e.g., when the information-processing load of the task becomes too high), children ignore the quantifiers "all" and "some" and treat all three versions of the question as generic. These result patterns lead to the intriguing implication that generics may be a kind of default for young children. When information-processing demands are high, children may ignore quantifiers and interpret a wide range of utterances as generic.

This interpretation remains speculative at the present time. In future research, it would be important to test the generics-as-default hypothesis more directly. For example, one could place older children under increased processing demands to determine if generic interpretations increase in such contexts. Alternative interpretations would also need to be tested more directly. For example, it is also possible that 3-year-olds are ignoring the form of the questions entirely and simply answering "yes" on the basis of how readily they can retrieve examples of the property being linked to the kind mentioned. One way to begin to address this question would be to examine 3-year-olds’ understanding with properties that are highly available in context but not true of the generic kind. For example, 3-year-olds could be shown a picture of a dozen glasses of chocolate milk and then asked, “Does milk have chocolate in it?” If their responses to this narrow-scope question are influenced simply by the availability of instances that match the predicate, then they should tend to answer “yes.” In contrast, if their responses indicate a true generic response, then they should give a response intermediate between “yes” and “no” (as they have done in Study 1).

It is also intriguing that 3-year-olds succeeded on the all–some posttest, which used a small set of concrete, available objects. The reasons why 3-year-olds have difficulty applying "all" and "some" to categories considered more broadly remain unknown at the present time. Perhaps the task poses insurmountable information-processing demands by requiring the child to hold in mind abstract sets and subsets (e.g., the set of girls and the subset of girls with curly hair) as well as the scope of the quantifier. A similar argument has been made with regard to young children’s understanding of basic–subordinate level inclusion relations (Johnson, Scott, & Mervis, 1997). Perhaps 3-year-old children would have performed better on all three forms of the question in the main task ("all," “some,” and generic) if we had provided pictures depicting accurate samples of the categories in question, such as a set of girls, some with curly hair and some with straight hair. More generally, an important task for future research will be to determine what underlies the developmental changes identified in this study.

A final point that deserves further investigation concerns children’s interpretation of generics with respect to properties (e.g., "Dogs are furry") versus with respect to behaviors (e.g., “Birds fly”). Although questions of both types were included in Study 1, most of the questions concerned properties rather than behaviors, and thus it was not possible to compare property types within this data set. In Study 2, both sorts of questions were included simply for the sake of greater generalizability of the results. Interestingly, we found that there was a tendency for actions to appear relatively more frequently for generics and for physical properties to appear relatively less frequently for generics. It may be that children consider behaviors to be relatively more central to determining what something is, compared to physical properties (see Shipley, 2000, for some evidence in favor of this possibility). This tentative proposal is also consistent with the earlier finding that preschool children judge behaviors to be more innately determined than physical features (Gelman & Wellman, 1991).

We conclude that preschool children readily grasp the semantics of generic sentences even though generic noun phrases are neither overtly marked in language (i.e., there is no single linguistic form for conveying them) and even though their meaning could be characterized as less clear-cut compared to the logical precision of "all" and "some." Taken together, these findings suggest that generic concepts are readily available in young children’s category representations. Just as children are capable of reasoning about everyday objects in their world, so too can they construct and reason about generalized kinds. In future research it will be important to explore the implications of this knowledge for children’s world knowledge and category representations.

References
Appendix

Study 1 Items

<table>
<thead>
<tr>
<th>Wide-scope questions</th>
<th>Narrow-scope questions</th>
<th>Irrelevant questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>*Do animals eat?</td>
<td>*Do bears have white fur?</td>
<td>*Are children made of feathers?</td>
</tr>
<tr>
<td>*Do elephants have trunks?</td>
<td>*Do books have color pictures?</td>
<td>*Do fish have branches?</td>
</tr>
<tr>
<td>*Are fires hot?</td>
<td>*Do clothes have zippers?</td>
<td>*Do fruits have gas tanks?</td>
</tr>
<tr>
<td>*Do frogs have eyes?</td>
<td>*Do dogs have brown spots?</td>
<td>*Do garages sing?</td>
</tr>
<tr>
<td>*Do ice creams melt in the sun?</td>
<td>*Do girls have curly hair?</td>
<td>*Do magazines blow bubbles?</td>
</tr>
<tr>
<td>*Do refrigerators have doors?</td>
<td>*Do shirts have stripes?</td>
<td>*Do monkeys have beaks?</td>
</tr>
<tr>
<td>Do alligators have mouths?</td>
<td>Do birds live in cages?</td>
<td>Do couches have windows?</td>
</tr>
<tr>
<td>Is candy sweet?</td>
<td>Do dresses have pockets?</td>
<td>Do pencils have noses?</td>
</tr>
<tr>
<td>Do cars have engines?</td>
<td>Are flowers yellow?</td>
<td>Do penguins have telephones?</td>
</tr>
<tr>
<td>Do giraffes have long necks?</td>
<td>Does milk have chocolate in it?</td>
<td>Do pigs fly?</td>
</tr>
<tr>
<td>Do hammers have handles?</td>
<td>Do people have blonde hair?</td>
<td>Do saws have toothaches?</td>
</tr>
<tr>
<td>Do plants grow?</td>
<td>Are tools made of wood?</td>
<td>Do zebras wear watches?</td>
</tr>
</tbody>
</table>

Note. Items are listed here in generic form only. All items were presented to 4-year-olds and adults. Items preceded by an asterisk were presented to 3-year-olds.