Quantified statements are recalled as generics: Evidence from preschool children and adults

Sarah-Jane Leslie [a,*], Susan A. Gelman [b]

[a] Department of Philosophy, 1879 Hall, Princeton University, Princeton, NJ 08544, United States
[b] Department of Psychology, 530 Church Street, University of Michigan – Ann Arbor, Ann Arbor, MI 48109, United States

ARTICLE INFO

ABSTRACT

Generics are sentences such as "ravens are black" and "tigers are striped", which express generalizations concerning kinds. Quantified statements such as "all tigers are striped" or "most ravens are black" also express generalizations, but unlike generics, they specify how many members of the kind have the property in question. Recently, some theorists have proposed that generics express cognitively fundamental/default generalizations, and that quantified statements in contrast express cognitively more sophisticated generalizations (Gelman, 2010; Leslie, 2008). If this hypothesis is correct, then quantified statements may be remembered as generics. This paper presents four studies with 136 preschool children and 118 adults, demonstrating that adults and preschoolers alike tend to recall quantified statements as generics, thus supporting the hypothesis that generics express cognitively default generalizations.

© 2011 Elsevier Inc. All rights reserved.

1. Introduction

Much of cognitive psychology has focused on how our core conceptual knowledge of the world is represented (e.g., Carey, 2009; Gelman, 2003; Holyoak & Morrison, 2005; Keil, 1989; Murphy, 2002; Rosch, 1973; Smith & Medin, 1981; and myriad others). A significant portion of this core conceptual knowledge consists of generalizations about *kinds* (e.g., Gelman, 2003; Margolis & Laurence, 1999; Prasada, 2000). For example, the beliefs that connect *tigers* with *stripes*, *dogs* with *tails*, and *doctors* with *healing people* all involve the generalization of properties to kinds. What is the nature of these kind-based generalizations? Are they at heart quantitative and statistical (e.g., Rosch, 1973), or some-

---

* Corresponding author.
  E-mail address: sjleslie@princeton.edu (S.-J. Leslie).

thing more complex and richly structured (e.g., Carey, 2009; Gelman, 2003; Prasada & Dillingham, 2006, 2009)?

Researchers have studied these questions using an enormously wide variety of methods. One example includes the body of empirical work devoted to studying the nature of kind-based generalizations in memory (e.g., McCarthy, 1995; Meyer, 1970; Smith, Shoben, & Rips, 1974). In particular, researchers have for many years studied the nature of semantic memory (or 'gist memory') as a way of assessing which aspects of a target will be encoded, and which will be forgotten (e.g., Jacoby & Dallas, 1981; Kintsch, 1988; Tulving, 1972). Thus, classic and ongoing studies of semantic memory provide a window onto aspects of conceptual representation.

This paper presents a series of studies conducted on both children and adults, designed to assess their semantic memory for different types of kind-based generalizations. Kind-based generalizations may be either quantificational (e.g., "All tigers have stripes") or generic (e.g., "Tigers have stripes"). Which (if either) of these forms corresponds to our core conceptual beliefs about kinds? Answering this question has important consequences for the study of concepts and categorization. If our most basic way of generalizing information about kinds is quantificational, this means that it can be characterized in quantitative, statistical terms – that our semantic knowledge is organized in terms of logic sets and set-inclusion relations, as these are the hallmarks of quantificational generalizations (e.g., Barwise & Cooper, 1981). Generic generalizations, in contrast, cannot be characterized in these terms, but rather reflect richer and more complex relations between the kind and the property, which cannot be reduced to purely formal, quantitative terms (Carlson, 1977; Leslie, 2007). The precise nature of generic generalizations is not yet fully understood (but see Cimpian, Brandone, & Gelman, 2010; Cimpian, Gelman, & Brandone, 2010; Khemlani, Leslie, & Glucksberg, 2009, in press; Leslie, 2008; Prasada & Dillingham, 2006, 2009 for some important work).

The main focus of this work is the nature of the underlying conceptual knowledge that is *expressed by* the linguistic form, rather than the linguistic form itself. However, understanding how our conceptually fundamental generalizations are expressed in language is itself of interest for several reasons. For example, it provides insight into subtle linguistic differences with implications for implicit messages that are conveyed in ordinary conversation. Further, because language is a primary source of information to children regarding regularities in the world around them (Gelman, 2010), this investigation can shed light on how conceptually important information is communicated to young children by parental speech—and how they are interpreted by young children as they hear such speech.

Recent theoretical and empirical work has begun to address whether general conceptual knowledge is quantificational or generic. In particular, Leslie (2007, 2008) and Gelman (2010) argue that generics, not quantified statements, articulate conceptually central generalizations. They argue that generics express cognitively fundamental, default generalizations, and that quantified statements, in contrast, express more sophisticated, less accessible generalizations. We briefly review the data in favor of this claim, and then report a series of experiments designed to further test the predictions of this hypothesis.

Children begin to produce generics around age two-and-a-half, and by the time they reach 3–4 years, they produce generics as frequently as adults (Gelman, Goetz, Sarnecka, & Flukes, 2008). Recent evidence suggests that two-and-a-half-year-olds understand that generics express general claim about kinds, rather than specific claims about individuals (Gelman & Raman, 2003; Graham, Nayer, & Gelman, 2011). Thus, as soon as children master the requisite syntactic skills, they produce and understand generics (Gelman, 2010; Gelman et al., 2008).

When it comes to expressing generalizations about kinds, generics are used much more frequently than universal quantifiers (e.g., "all", "every", "any") in both children's speech and in child-directed speech (Gelman, 2003; Gelman, Hollander, Star, & Heyman, 2000). Universal quantifiers are used primarily in reference to specific sets of individuals (e.g., "pick up all those toys") rather than to kinds as a whole (e.g., "all tigers have stripes"). Hollander, Gelman, and Star (2002) found that both 3- and 4-year-olds could appropriately interpret the quantifiers "all" and "some" in reference to a specific set of items (e.g., "are all the crayons in the box?"); however, they also found that 3-year-olds had difficulty with "all" and "some" when the quantifiers were used to make kind-wide generalizations (e.g., "are all fires hot?"). Interestingly, while Hollander et al. found significant developmental differences in the case of category-wide uses of "all" and "some", they found no developmental differences in the

case of generics, suggesting that generics may be easier for young children to understand than quantifiers. Tardif, Gelman, Fu, and Zhu (2011) recently replicated these experiments with Mandarin speaking preschoolers, and found that Mandarin speaking 3- and 4-year-olds had difficulty with category-wide uses of quantifiers.

Hollander et al. (2002) and Tardif et al. (2011) further found that the pattern of errors that English- and Mandarin-speaking preschoolers made with "all" and "some" specifically involved their responding to quantified statements as though they were generics. That is, English- and Mandarin-speaking preschoolers gave the same pattern of responses regardless of whether they were presented with quantified statements (using "all" or "some") or with generics, and further, this pattern was indistinguishable from the pattern of responses that older children and adults gave when presented with generics. Thus, these results may potentially reflect that the preschoolers relied on their interpretations of generics when confronted with category-wide statements quantified with "all" and "some".

Might these results reflect a more general tendency to assimilate quantified statements to generics? Leslie, Khemlani, and Glucksberg (2011) report evidence suggesting that adults may exhibit this tendency under some circumstances. In particular, adults sometimes accept statements such as "all ducks lay eggs", despite knowing that male ducks don't lay eggs, and even despite having explicitly judged minutes before that male ducks don't lay eggs (Experiment 3). This finding would be explicable if English-speaking adults, like English- and Mandarin-speaking preschoolers, were relying on their interpretation of the widely accepted generic "ducks lay eggs", instead of evaluating the incorrect universal "all ducks lay eggs" (Khemlani, Leslie, Glucksberg, & Rubio-Fernandez, 2007; Leslie et al., 2011; see also Jönsson & Hampton, 2006; Sloman, 1993, 1998, for results that are readily explicable on this hypothesis).

From a theoretical perspective, the question remains *why* children and adults have this tendency. Leslie (2008) and Gelman (2010) hypothesize that generics express cognitively default generalizations, whereas quantified statements express more taxing and cognitively sophisticated ones. Such a hypothesis would explain why generics are so easy for children to acquire, and why they are the preferred way of expressing kind-based generalizations both in children's speech and in child-directed speech: generic generalizations are the easiest sort of generalization for young children to process. It would also explain the tendency to interpret quantified statements as though they were generics. Because quantified statements involve more taxing, non-default generalizations, people sometimes fall back on the default generic interpretation instead of processing the more taxing one.

Leslie (2008) also notes that such a generics-as-default hypothesis would explain an otherwise puzzling cross-linguistic fact. To express a quantified generalization in English, we have to use a specific quantifier – for example, "*all*(/*most*/*some*) tigers are striped". To express a generic generalization, however, we just 'omit' the quantifier: "tigers are striped". There is no special linguistic marker ("gen") that specifically signals a generic. Interestingly, this is not an isolated fact about English: no known language has a word or linguistic marker that specifically signals a generic (Dahl, 1985; Krifka et al., 1995). In this way, generics are less syntactically marked than quantified statements. Chomsky (2000) argues that less marked syntactic forms may be associated with default interpretations. Thus, if generics express cognitively default generalizations, this would explain why they are syntactically less marked than quantified statements.

The present paper reports a series of experiments designed to further test the generics-as-default hypothesis. The hypothesis was investigated using a memory task modeled on Gelman and Raman (2007), in which participants were presented with a series of pictures of animals, and told a novel fact about each kind of animal. Following a distractor task, they were asked to recall the fact they had been taught, upon being presented with the picture of the relevant animal. Gelman and Raman (2007) used this paradigm to test children's and adults' memory for generic vs. specific statements (e.g., "cats sweat through their paws" vs. "this cat sweats through its paws"). They were especially interested in memory for the type of sentence – generic vs. specific – and found that 3- and 4-year-olds and adults all tended to successfully recall generics as generics, and specifics as specifics.

Gelman and Raman (2007) were primarily interested in how well young children differentiated information about individuals from information about kinds. We sought to adapt their paradigm to test a different set of predictions: namely that both children and adults would recall quantified statements as generics. If generics express cognitively more default generalizations than quantified state-

ments do (Gelman, 2010; Leslie, 2008), then quantified statements may be more likely to be stored in memory as generics than vice versa.

Previous tests of the generics-as-default hypothesis have relied on participants' evaluations of quantified statements, to see if there is evidence that they sometimes substitute their judgments of generics for their judgments of quantified statements (e.g., Hollander et al., 2002; Khemlani et al., 2007; Leslie et al., 2011; Tardif et al., 2011). (Conversely, some prior research has examined the recall of nouns vs. modifiers (e.g., Ehrlich, 1975), but to our knowledge, no one has focused on quantifiers and how they compare to generics.) The present experiments provide converging evidence by use of a distinctly different set of methods, which tap into a different set of behavioral responses.

There are numerous studies across many areas that examine errors in recall in order to assess cognitive defaults. One early example is found in psycholinguistics, in particular the study of the differential processing of negative vs. affirmative statements. Adults are more likely to incorrectly recall a negative statement (e.g., "John did not sing") as an affirmative statement (e.g., "John sang") than vice versa, which researchers took to indicate that the affirmative form is more basic than the negative form (e.g., Cornish & Wason, 1970; Fillenbaum, 1966; Mehler, 1963). Other diverse examples of this methodology include work on scripts, indicating that people have default expectations concerning how events should unfold, which give rise to errors in recall and recognition tasks (e.g., Bower, Black, & Turner, 1979), and work on the processing of stereotype consistent/inconsistent information, which indicates that stereotypical beliefs are, by default, applied to individuals, at least under high processing loads (e.g., Macrae, Hewstone, & Griffiths, 1993). Thus, the use of a memory paradigm would seem a promising way of testing the generics-as-default hypothesis.

In Experiment 1A, adult participants were presented with photos of animals, and told a novel fact about each animal. Half of the facts were presented in generic form, and the other half were presented as containing a particular quantifier – either "all", "most", or "some". We predicted, in accordance with the generics-as-default hypothesis, that participants would be more likely to recall quantified statements as generics than vice versa. Experiment 1B was designed to provide an even stronger test of the hypothesis that quantified statements will be recalled as generics. In Experiment 1B, adult participants never heard any generics in the learning phase – only quantified statements. We predicted that participants would nonetheless recall some of the statements as generics.

Experiments 2A and 2B followed the same design as Experiments 1A and 1B respectively, but were conducted with 3- and 4-year-old participants. Experiment 3 was designed in part to rule out a potential alternative explanation of 3-year-olds' performance, namely that they are simply forgetting the first word, i.e. the quantifier. That is, if 3-year-olds recalled all three quantified statements as generics, then this may just be due to a 'dumb' strategy of shortening the statement to exclude the first word. Experiment 3 addressed this by pairing generics with the negative quantifier "no" (e.g., "No spiders have ten legs"). If 3-year-olds are simply dropping the first word, then they should recall such statements as positive generics (e.g., "no spiders have ten legs" would be recalled as "spiders have ten legs"). Finally, Experiments 4A and 4B were designed to rule out additional alternative explanations for the results of the earlier studies.

2. Experiment 1A

Experiment 1A was designed to test adults' memory for generics vs. quantified statements. The design was adapted from Gelman and Raman (2007): participants were shown a series of photographs of different animals, and were told a novel fact about each animal. Participants heard half of the facts as generic statements, and the other half as statements that contain a particular quantifier (either "all", "most" or "some"). Each participant heard only one particular quantifier throughout. We predicted that quantified statements would be recalled as generic more often than generic statements would be recalled as quantified.

One outstanding question is whether all three quantifiers would be recalled as generic, or whether they might pattern differently. We predicted, following Leslie et al. (2011), that adults would have a tendency to recall statements quantified with "all" as generics, but what about "most" and "some"? Although several experiments have shown that adults differentiate "most" from generics (e.g., Cim-

Table 1
Items used in Experiments 1 and 2.

Category	Generic form	Quantified form
Bear	Bears climb trees	All/Most/Some bears climb trees
Rabbit	Rabbits keep food in their cheeks	All/Most/Some rabbits keep food in their cheeks
Snail	Snails breathe through their feet	All/Most/Some snails breathe through their feet
Hippo	Hippos like to swim	All/Most/Some hippos like to swim
Spider	Spiders shed their skin	All/Most/Some spiders shed their skin
Bee	Bees have five eyes	All/Most/Some bees have five eyes
Rhino	Rhinos have horns made out of hair	All/Most/Some rhinos have horns made out of hair
Camel	Camels store fat in their humps	All/Most/Some camels store fat in their humps
Seahorse	Seahorses come out of their dads' tummies	All/Most/Some seahorses come out of their dad's tummies
Cat	Cats sweat through their paws	All/Most/Some cats sweat through their paws

pian, Brandone, et al., 2010; Cimpian, Gelman, et al., 2010), no experiments have yet been conducted to test whether "most" statements are ever interpreted as generics. We predicted that "most" statements would also tend to be recalled as generics, since "most" conveys information that applies broadly to the members of the kind. The quantifier "some" may be interpreted somewhat differently by adults, however. Leslie et al. (2011) failed to find any evidence that adults ever treat "some" statements as generics. This may be because adults interpret "some" as conveying information about only a limited number of members of a category (Bott & Noveck, 2004), and so in this way "some" statements may be interpreted more like specific statements about a particular member or set of members of the kind. Thus we predicted that "some" statements would be more readily distinguished from generics in memory than would "all" and "most" statements.

2.1. Method

2.1.1. Participants

Forty-eight undergraduates (32 women; 16 men) were recruited from a large Midwestern public university. One additional participant was tested but dropped from analysis because of insufficient proficiency in English.

2.1.2. Materials

Ten photographs of animals were used.[1] Each animal had a novel fact associated with it. The photographs and the facts were a subset of those used in Gelman and Raman (2007); see Table 1. Gelman and Raman (2007) pretested these items with preschoolers, and found that the facts were unfamiliar to children. Each fact was either presented in generic form or in quantified form ("all", "most", or "some").

2.1.3. Procedure

Each participant was tested individually. They were asked to listen carefully, because they would be asked to remember the information provided. They were presented with five test items, one at a time. For each item, the participant first saw a photograph depicting the animal, and then heard a novel fact associated with the animal. For example, participants saw a photo of a bear and heard one of the following sentences: "Bears climb trees" (generic), "All bears climb trees" (all), "Most bears climb trees" (most), or "Some bears climb trees" (some). Following presentation of the first five items, participants received a 4-min distractor task in which they were asked to solve a series of multidigit multiplication problems. They were then tested for their recall of the five sentences. For each, they were shown the same photograph as they had seen during the initial presentation phase, and asked to recall what they had been told about the animal depicted. The entire process was then repeated with the

---

[1] Participants in Studies 1A, 1B, 2A, and 2B each received an additional six sentences (therefore 16 total). However, due to experimenter error, the additional sentences were not counterbalanced across conditions. We therefore report only those results from the 10 sentences that were properly counterbalanced. The major patterns of results presented here are equivalent to those from the full set of 16 items.

remaining five items. Responses were audiotaped and transcribed. The testing process took approximately 12 min to complete.

Participants were randomly assigned to one of three groups: generic/all, generic/most, and generic/some. Each group heard half the facts in generic form, and the other half in quantified form (either "all", "most", or "some", depending on the condition). Assignment of item to wording condition (i.e., generic vs. quantified) was counterbalanced. Order of items within each block of 5 was presented in one of two random orders.

2.1.4. Coding

The transcribed responses were independently coded by two researchers. Responses were coded as Generic, Some, Most, All, or Other. The Generic category included indefinite singular generics (e.g., "a tiger is striped") but excluded definite singular generics (e.g., "the tiger is striped"), because the latter is ambiguous between a generic and a specific reading. We also adopted the conservative practice of coding any statements involving adverbs of quantification (e.g., "always", "mostly", "usually", "sometimes") as belonging to the semantically closest quantifier category (e.g., "always" as All, "mostly" and "usually" as Most, "sometimes" as Some). This is consistent with standard linguistic treatments of these adverbs (e.g., de Swart, 1993; Heim, 1982; Kratzer, 1995; Lewis, 1975). Any other statements, including a failure to recall anything, were coded as Other. Partial recall of a sentence was coded, as were multiple forms in a given response (e.g., "cats sweat through their paws…. Most cats sweat through their paws" was coded as both Generic and Most). Agreement on the different coding categories ranged from 99% to 100%, and Cohen's Kappas ranged from .98 to 1.00. All discrepancies in coding were resolved by discussion.

2.2. Results

2.2.1. Percentage correctly recalled

We calculated the percentage of trials that were recalled in the correct format (generic, "all", "some", or "most", depending on the wording of the sentence that had been presented). These scores were entered into a 2 (sentence type: generic, quantified) × 3 (condition: generic/all, generic/most, generic/some) ANOVA, with condition as the between-subjects factor and sentence type as the within-subjects factor.

We obtained a main effect of sentence type, $F(1, 45) = 13.55$, $p = .001$, $\eta^2 = .23$, indicating that generic sentences were recalled correctly more often than quantifiers (67.5% vs. 45%, respectively). We also obtained a main effect of condition, $F(2, 45) = 3.56$, $p < .05$, $\eta^2 = .14$, indicating that recall was higher in the generic/some condition than in either of the other two conditions. (All η^2 results that we report use the partial η^2 formula (SSeffect/(SSeffect + SSerror)). Tabachnick and Fidell (1989) suggest that partial η^2 is an appropriate alternate computation of η^2.)

However, this main effect must be interpreted in light of the sentence type × condition interaction, $F(2, 45) = 5.74$, $p < .01$, $\eta^2 = .20$. For the generic/all and generic/most conditions, generic sentences tended to be recalled more accurately than quantified sentences, $p = .10$ and $p < .001$, respectively (see Fig. 1). In contrast, for the generic/some condition, generic sentences and quantified sentences were correctly recalled at equal rates.

2.2.2. Percentage recalled as opposite (generic as quantified and vice versa)

We calculated the proportion of trials that were recalled in the opposite form to what was presented (generic sentences recalled as quantified; quantified sentences recalled as generic). These scores were entered into a 2 (sentence type: generic, quantified) × 3 (condition: generic/all, generic/most, generic/some) ANOVA, with condition as the between-subjects factor and sentence type as the within-subjects factor.

We obtained a main effect of sentence type, $F(1, 45) = 9.38$, $p < .01$, $\eta^2 = .17$, indicating that quantified sentences were more often recalled as generic (38% of trials) than generic sentences were recalled as quantified (18%). We also obtained a main effect of condition, $F(2, 45) = 4.03$, $p < .05$, $\eta^2 = .16$. Overall, participants more often recalled sentences in the opposite form in the generic/all and generic/most conditions than in the generic/some condition. However, this finding must be interpreted in light of
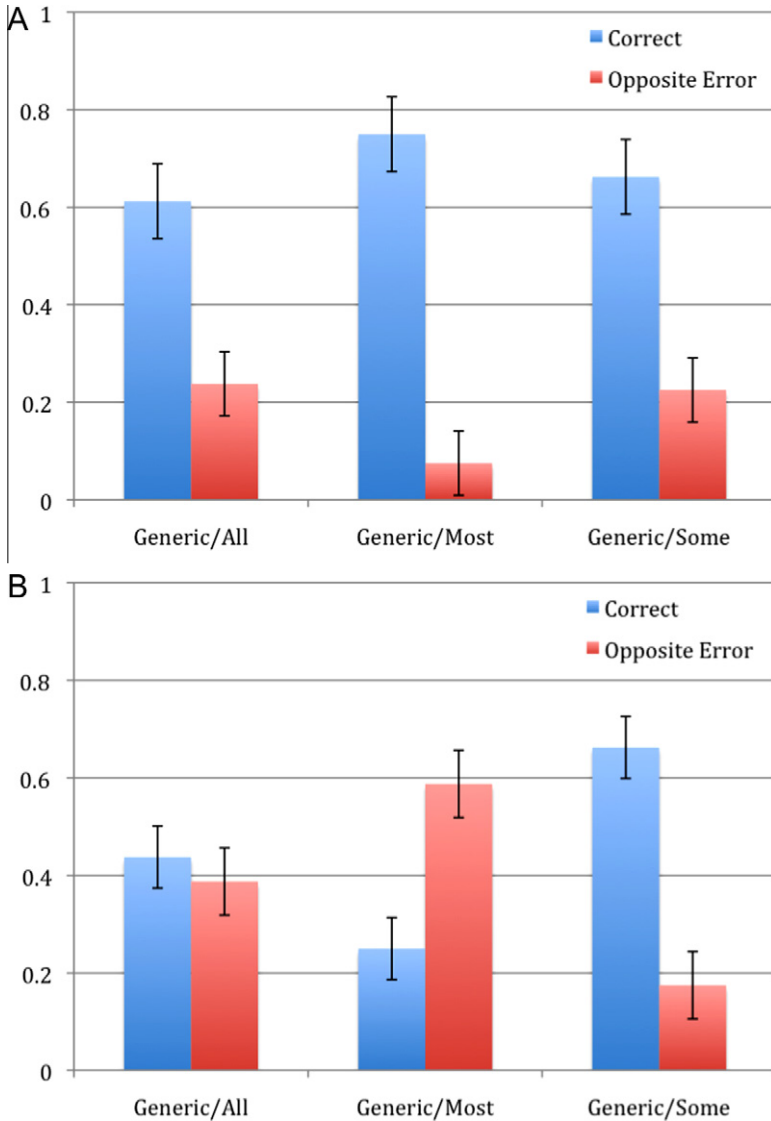
Fig. 1. (A) Experiment 1A, mean proportion of generic sentences recalled in correct vs. opposite format. (B) Experiment 1A, mean proportion of quantified sentences recalled in correct vs. opposite format.

the significant sentence type × condition interaction, $F(2,45) = 6.10$, $p = .005$, $\eta^2 = .21$. This interaction shows that, for the generic/most condition, quantifiers are more often recalled as generic than vice versa (59% vs. 7%), $p < .001$. For the generic/all condition, there was a non-significant trend towards the quantifier being recalled more often as generic than vice versa (39% vs. 24%), $p = .20$. In contrast, for the generic/some condition, there is no difference between how often quantifiers are recalled as generic and generics are recalled as quantified (22% vs. 17%), $p > .6$.

2.2.3. Correct responses vs. opposite responses

We next were interested in the relative frequency of correct vs. opposite responses, to assess the strength of the tendency (predicted by the generics-as-default hypothesis) for adults to recall quanti-

fiers as generic. We therefore conducted a series of paired *t*-tests, comparing the percentage of correct responses with the percentage of opposite responses within each condition, as a function of sentence type. The results appear in Fig. 1. For the generic/all condition, generic sentences were recalled correctly more often than in opposite (all) form, $t(15) = 2.47$, $p = .026$, whereas quantified sentences were recalled equally often as generic (opposite response) and as "all" (correct response), $t(15) = 0.38$, $p > .7$. For the generic/most condition, generic sentences were recalled correctly more often than they were recalled as "most" (opposite response), $t(15) = 6.45$, $p < .001$, whereas quantified sentences were recalled as generic (opposite) more often than they were recalled as "most" (correct), $t(15) = -2.30$, $p = .036$. Finally, for the generic/some condition, sentences were recalled in correct format more often than in opposite format, regardless of whether the sentence being recalled had been presented in generic or quantified form, $ts(15) > 3.1$, $ps < .01$. Altogether, these data reveal an asymmetry between generics and the quantifiers "all" and "most", but no asymmetry between generics and the quantifier "some".

2.2.4. Percentage of errors recalled as opposite

The analysis of "opposite" errors could reflect overall accuracy rates. In other words, if participants are more accurate in recalling generics, then the absolute rate of opposite errors will necessarily be relatively low. Therefore, we also examined the opposite errors as a percentage of errors. In other words, we ask whether there are systematic differences in the kinds of recall errors, controlling for differential rates of accuracy. To conduct this analysis, we gave each participant scores for the percentage of errors that showed the "opposite" pattern (generics recalled as quantified; quantifiers recalled as generic), separately for each wording type. Whenever a participant had no errors, their scores were not included in the analysis (remaining *N*s were 12 for the generic/all condition, 10 for the generic/most condition, and 8 for the generic/some condition). We then conducted a 2 (sentence type: generic, quantified) × 3 (condition: generic/all, generic/most, generic/some) ANOVA, with condition as the between-subjects factor and sentence type as the within-subjects factor. This analysis revealed no significant effects for sentence type or condition, although in the generic/most condition, quantified sentences were recalled as generic (59%) nearly twice as often as generic sentences were recalled as quantified (30%), $p = .08$, one-tailed.

2.3. Discussion

As predicted, generics tended to be correctly recalled as generics, whereas quantified statements were frequently recalled as generics. Participants showed little tendency to recall generics as quantified statements, but showed a marked tendency to recall quantified statements as generics. This effect was driven by "most" and to a lesser extent "all"; "some" statements were no more likely to be recalled as generics than generics were to be recalled as "some" statements. This finding is consistent with Leslie et al.'s (2011) findings: whereas adults sometimes interpreted "all" statements as generics in their tasks, they showed no tendency to interpret "some" statements as generics. One outstanding question from Leslie et al. is whether this tendency is specific to the quantifier "all", or whether it is the quantifier "some" that is exceptional in this regard. Our data show that "most" statements are recalled as generics, which suggests that the general phenomenon is not limited to the quantifier "all". Indeed, the tendency for quantifiers to be recalled as generic was strongest for "most", and weaker (though present) for "all". A possible explanation for why "some" is treated differently is that adults understand "some" statements to express distinctly limited generalizations. That is, adults may interpret "some" statements as pertaining only to a limited subset of the kind – not as a broad generalization over members of the kind like "all" and "most" statements – and generics may only be default interpretations for generalizations that pertain broadly to the members of the kind. We return to this possibility in the discussion of Experiment 2A. First, however, we investigated whether adults would recall quantified statements as generics even if they did not hear any generics in the learning phase.

3. Experiment 1B

In Experiment 1A, each participant heard half of their statements in generic form. To what extent does the tendency to recall quantified statements as generics depend on participants *hearing* generics? Clearly the results of Experiment 1A cannot be explained by simple confusion between the two types of statements heard, because in the case of "all" and "most", the tendency to recall quantified statements as generics outstripped the tendency to recall generics as quantified statements. However, an even stronger test of the predictions of the generics-as-default hypothesis would be to see if participants recall generics even if they only hear quantified statements in the learning phase.

In Experiment 1A, not a single participant recalled a statement as quantified with a quantifier *other* than the one they heard in the learning phase. Errors overwhelmingly consisted of participants either making 'opposite' errors, or failing to remember anything at all. Would this still be the case if they heard no generics, however? We predicted that the tendency to recall quantified statements as generics would be so strong that, even if generics were not included in the learning phase, participants would still produce them in the test phase.

Experiment 1B tested this prediction by pairing "most" with "some". We selected these quantifiers for two reasons. "Most" was selected because participants had displayed the strongest tendency to recall "most" statements as generics. One possible alternative explanation for this might be that "most" statements are for some reason especially hard to remember, and easily confused with other statements. If this explanation was correct, then one would expect participants often to recall "most" statements as "some" statements in this task, so that the rate of opposite errors in this task would be comparable to the rate of opposite errors in the "most" vs. generics task in 1A. If, however, the generics-as-default interpretation is correct, then there should be fewer opposite errors with "most" in this experiment than in 1A. The quantifier "some" was chosen because the structure of 1B allowed for a more subtle test of whether adults have any tendency to recall "some" statements as generics. If there is no such tendency, then "some" statements here should not be recalled as generics – the primary errors with "some" would be opposite errors.

3.1. Method

3.1.1. Participants

Eighteen adults (3 women, 15 men) were recruited from a large Midwestern university. None had participated in the previous study.

3.1.2. Materials

The same materials as in Experiment 1A were used.

3.1.3. Procedure

The procedure was the same as in Experiment 1A, except that the participants all heard half the statements as "some" statements, and the other half as "most" statements.

3.1.4. Coding

The same coding practice as in 1A was adopted. Agreement was 100%.

3.2. Results

We conducted a series of paired *t*-tests to determine whether there were significant differences in recall for sentences presented with 'most' vs. 'some'. On all three measures, there were no significant differences in recall for 'most' vs. 'some': percent correct (33% for "most"; 42% for "some"), percent recalled as opposite (23% for most recalled as "some"; 18% for some recalled as "most"), and percent of errors recalled as opposite (31% of most recalled as "some"; 23% for some recalled as "most"), ps > .40.

We were interested in whether the errors with "most" found in Experiment 1A were due to a specific tendency to recall "most" as generics, or whether "most" just tends to be recalled as whatever other type of sentence it is paired with. To test this, we compared the rate of opposite errors with "most" in Experiment 1A (59% of trials) with the rate of opposite errors with "most" in this experiment (23%). More opposite errors occurred with "most" in Experiment 1A, $t(32) = 3.45$, $p = .002$, suggesting that the high error rate in Experiment 1A is not due to "most" being somehow unmemorable. Opposite errors with "some" here occurred at the same low rate as in 1A: 17.5% in 1A and 18% in 1B, $t(32) = .04$, $p > .9$.

The primary purpose of this experiment was to examine whether quantified sentences are ever recalled as generic. In contrast to Experiment 1A, where errors only entailed forms that had been presented in the original sentence block, here we found that participants consistently recalled quantified forms as generic. Over 60% of participants recalled a quantified sentence as generic as least once ($N = 11$ out of 18 participants). They did so more often for "most" sentences (21%) than "some" sentences (12%), t-paired(17) = 2.12, $p < .05$. Notably, errors in which the quantified sentences were recalled as generic were as frequent as errors in which the quantified sentences were recalled in the "opposite" quantified form, for both 'most' and 'some' sentences, $ps > .30$.

Finally, as in Experiment 1A, we examined the relative frequency of correct vs. opposite errors. We predicted that the overwhelming tendency in Experiment 1A to produce opposite errors in the "most" condition would no longer be found, given that the opposite wording in this case is another quantifier rather than a generic. And indeed this was the case: correct responses (33%) and opposite errors (23%) occurred at equal rates, $p > .3$. Furthermore, as in Experiment 1A, in the "some" condition, correct responses (42%) were more frequent than opposite errors (18%), $p = .039$.

3.3. Discussion

Participants sometimes recalled statements as generics, despite never hearing a generic in the learning phase. As in 1A, participants never recalled any statement as containing a quantifier that they did not hear in the learning phase. It is thus remarkable that they recalled generics, despite only hearing quantified statements in the learning phase. Even though participants did not show a bias towards recalling generics in the case of "some" statements in 1A, in 1B they recalled "some" statements as generics on 12% of trials. More specifically, "some" statements were recalled as generics just as frequently as they were recalled as "most" statements – i.e. the rate at which "some" statements were recalled as generics was equivalent to the rate of opposite errors for "some". Further, the majority of participants recalled at least one "some" statement as a generic. These findings suggest that even for "some" statements, there may be some underlying bias towards assimilating them to generics.

Statements quantified with "most' were recalled as generics here more frequently than "some" statements were recalled as generics, which is as one would expect given the strong tendency to recall "most" as generics in 1A. Further, the rate of opposite errors for "most" here was significantly lower than in 1A. If "most" had been recalled as "some" here as frequently as "most" statements were recalled as generic in 1A, this would have suggested that "most" statements were just easily confused with other statements, and would have undermined the interest of the findings in 1A. However, the rate of opposite errors in this study was significantly lower than in 1A. This finding, coupled with the substantial rate at which "most" statements were recalled as generics here, reflects that these results are due to a sizeable underlying tendency to remember "most" statements as generics.

Experiments 1A and 1B thus show that adults have a distinct tendency to recall statements quantified with "all" and "most" as generics, and also have a smaller tendency to recall "some" statements as generics, as predicted by the generics-as-default hypothesis. In Experiments 2A and 2B, we extended our findings to 3- and 4-year-old children.

4. Experiment 2A

Experiment 2A followed the same design as Experiment 1A, but was conducted with 3- and 4-year-old participants. The generics-as-default hypothesis predicts that, like adults, preschool children

should recall quantified statements as generics. The tendency to do so should be at least as strong among young children as it is among adults.

Furthermore, we examine potential differences among the quantifiers "all", "most", and "some". For adult participants, the tendency to recall "some" statements as generics is weak compared to "all" and "most" statements. Why might this be? One explanation could be that adults interpret "some" as expressing a very bounded and restricted sort of generalization – they understand it to mean that *only a few* members of the kind have the property. Understanding "some" this way entails showing a scalar implicature effect (Bott & Noveck, 2004) – that is, believing that "some" implies "not all". Developmental evidence can speak to this issue, because very young children do not show a scalar implicature effect (Barner, Chow, & Yang, 2009; Huang & Snedeker, 2009; Hurewitz, Papafragou, Gleitman, & Gelman, 2006; Noveck, 2001; Papafragou & Musolino, 2003). So it is possible that for the youngest children, "some" expresses an open and unbounded generalization which is more easily assimilated to a generic generalization. Later in development, "some" comes to be understood as expressing a more bounded and restricted generalization, and in this way is more similar to a specific claim than to a generic generalization. Consistent with this interpretation, Hollander et al.'s (2002) and Tardif et al.'s (2011) data suggest that English speaking 3-year-olds, but not 4-year-olds, may interpret "some" statements as generics. If this is part of a more general tendency, then one would likely find that 3-year-olds recall "some" statements, as well as "all" and "most" statements, as generics.

4.1. Method

4.1.1. Participants

Twenty-six 3-year-olds (12 girls, 14 boys; mean age = 3.67 years; range = 3.04–3.99 years) and thirty 4-year-olds (14 girls, 16 boys; mean age = 4.53 years; range = 4.04–5.21 years) were recruited from preschools in a small Midwestern city. Sixteen children were assigned to the generic/all condition, 16 to the generic/most condition, and 24 to the generic/some condition. More children were assigned to the generic/some condition so that we could examine more closely the effects of age within this condition. An additional 10 participants (four 4-year-olds and six 3-year-olds) were dropped from the study because they did not recall at least one predicate from each wording condition.

4.1.2. Materials

The same materials as in Experiment 1 were used.

4.1.3. Procedure

The main procedure was the same as in Experiment 1A, except that the children played with Legos as a distractor task.

The main phase was then followed by a post-test, in which each child was asked to respond 'yes' or 'no' to three questions. The questions involved whatever quantifier the child had heard in the first phase. For the first question, two crayons were place in a box, while four were left outside the box, and the child was asked, "Are all/most/some crayons in the box?" Then two more crayons were moved into the box, and the child was asked the same question again. Finally, the remaining two crayons were moved into the box, and the child was asked the question a third time.

4.1.4. Coding

The same coding practice as in Experiment 1 was used. Agreement across the different coding categories ranged from 95% to 100%, and Cohen's Kappas ranged from .885 to 1.00.

4.2. Results

4.2.1. Post-test

Children performed extremely well on "all", correctly denying that 2/6 or 4/6 are "all" (100%), and correctly reporting that 6/6 is "all" (93%). Children also performed well on "some", correctly reporting that 2/6 is "some" (73% at age 3, 100% at age 4) and that 4/6 is "some" (91% at age 3, 100% at age 4).

There was also sharp developmental change in how often children treated "some" as contrastive with "all" (i.e., the scalar implicature): 9% of 3-year-olds and 60% of 4-year-olds did so, which is consistent with Barner et al. (2009), Hurewitz et al. (2006), and Pouscoulous, Noveck, Politzer, and Bastide (2007), whose results taken jointly suggest that 4-year-olds sometimes show a weak scalar implicature effect, whereas 3-year-olds tend not to. Finally, children performed poorly on "most" (only 42% correctly reporting that 2/6 is not "most," and only 67% correctly reporting that 4/6 is "most"), which is consistent with previous work on children's comprehension of "most" (Barner et al., 2009; Halberda, Taing, & Lidz, 2008; Papafragou & Schwarz, 2005/2006). Thus results for "most" should be interpreted cautiously. The majority of children (83%) did not consider "most" to be contrastive with "all" (replicating Papafragou & Schwarz, 2005/2006), although the children who did so were both 4-year-olds.

For the primary analyses, we first collapsed across age groups (3-year-olds, 4-year-olds), examining the percent correct recall, the percent opposite recall, and the percent of errors that were recalled as opposite.

4.2.2. Analyses collapsing across age groups

4.2.2.1. Percentage correct recall.
We calculated the proportion of trials that were recalled in the correct format (generic, "all", "some", or "most", depending on the wording of the sentence that had been presented). These scores were entered into a 2 (sentence type: generic, quantified) × 3 (condition: generic/all, generic/most, generic/some) ANOVA, with condition as the between-subjects factor and sentence type as the within-subjects factor. See Fig. 2 for results.

We obtained a main effect of sentence type, $F(1,53) = 17.28$, $p < .001$, $\eta^2 = .25$. Recall was higher for generic sentences (41%) than for quantified sentences (18%). Although there was no significant interaction with condition, we conducted post-hoc tests to determine whether the effect of sentence type held up within each condition considered separately (generic/all, generic/most, and generic/some), and found that generics were recalled better than quantifiers within each, $ps < .05$ for generic/all and generic/most conditions; $p = .059$ for generic/some.

4.2.2.2. Percentage recalled as opposite (generic as quantified and vice versa).
We calculated the proportion of trials that were recalled in the opposite form to what was presented (generic sentences recalled as quantified; quantified sentences recalled as generic). These scores were entered into a 2 (sentence type: generic, quantified) × 3 (condition: generic/all, generic/most, generic/some) ANOVA, with condition as the between-subjects factor and sentence type as the within-subjects factor. We obtained a main effect of sentence type, $F(1,53) = 20.24$, $p < .001$, $\eta^2 = .28$, indicating that quantifiers were more often recalled as generics (39%) than generics were recalled as quantified (12%). This result held up within each condition examined individually, all $ps < .05$.

4.2.2.3. Correct responses vs. opposite responses.
We next were interested in the relative frequency of correct vs. opposite responses, to assess the strength of the tendency (predicted by the generics-as-default hypothesis) to recall quantifiers as generic. We therefore conducted a series of paired t-tests, comparing the percentage of correct responses with the percentage of opposite responses within each condition, as a function of sentence type. For generics, correct responses always exceeded opposite (quantified) errors, all $ps < .05$. In contrast, for quantifiers, correct responses never exceeded opposite (quantified) errors. For both "all" and "most", opposite (generic) errors exceeded correct responses ($ps = .01$ and .015, respectively). For "some", opposite errors were equivalent in frequency to correct responses, $p > .2$.

4.2.2.4. Percentage of errors recalled as opposite.
As in Study 1A, we also examined the opposite errors as a proportion of errors, in order to control for differential rates of accuracy across the different cells of the design. To conduct this analysis, we gave each participant scores for the percentage of errors that showed the "opposite" pattern (generics recalled as quantified; quantifiers recalled as generic), separately for each wording type. Three participants in the generic/most condition had no errors; their scores were not included in the analysis. We then conducted a 2 (sentence type: generic, quantified) × 3 (condition: generic/all, generic/most, generic/some) ANOVA, with condition as the between-subjects factor and sentence type as the within-subjects factor.
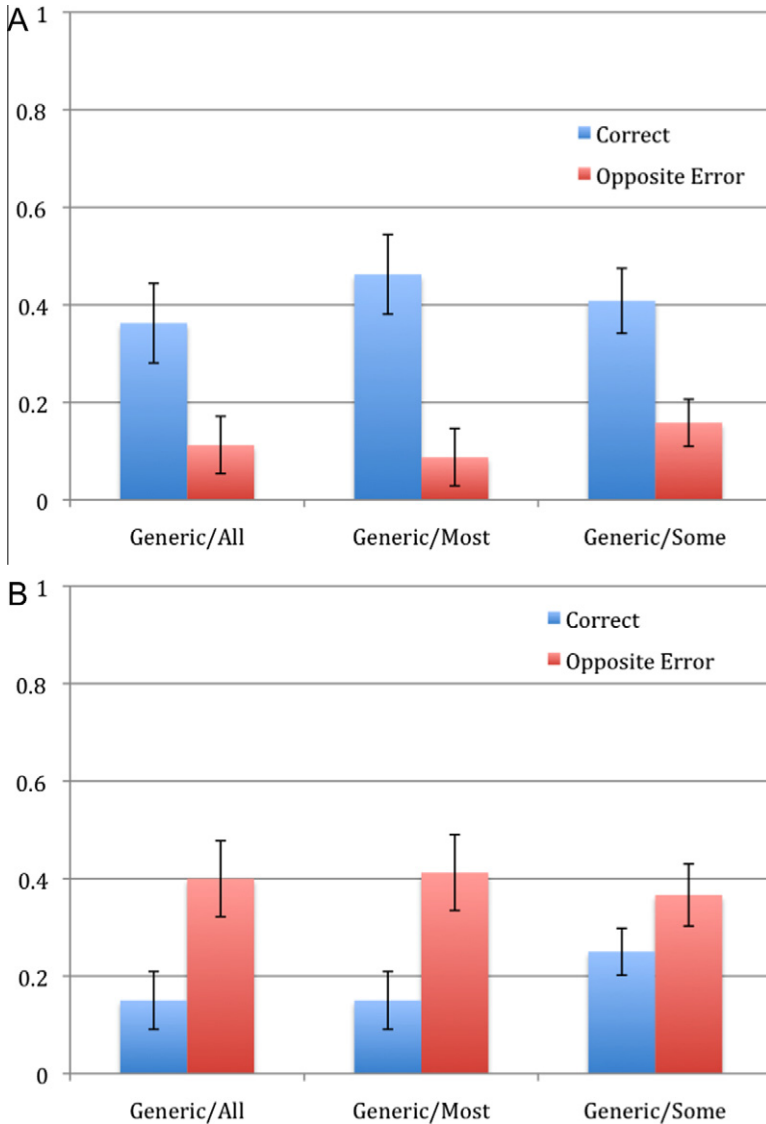
Fig. 2. (A) Experiment 2A, mean proportion of generic sentences recalled in correct vs. opposite format. (B) Experiment 2A, mean proportion of quantified sentences recalled in correct vs. opposite format.

We obtained a main effect of sentence type, $F(1,48) = 17.82$, $p < .001$, $\eta^2 = .27$, indicating that, even controlling for error frequency, children recalled quantifiers as generic more frequently (46% of errors) than they recalled generics as quantified (18% of errors). Post-hoc comparisons indicated that this distinction held up within each of the conditions separately (generic/all, generic/most, generic/some), ps < .05.

4.2.3. Analyses separated by age group

In order to examine developmental differences between 3- and 4-year-olds, we next conducted a set of analyses that separates the two age groups. We were particularly interested in how performance

compared across ages in the generic/some condition, given the age differences in performance on the post-test with "some" (with greater evidence for scalar implicatures at age 4 than age 3). We therefore collapsed the data from generic/all and generic/most into a single group.

We calculated the proportion of trials that were recalled in the correct format, and entered these scores into a 2 (sentence type: generic, quantified) × 2 (condition: generic/all or most, generic/some) × 2 (age: 3 years, 4 years) ANOVA, with condition and age as between-subjects factors and sentence type as the within-subjects factor. Here we report only those effects involving age, as that was the only novel factor. There was a main effect for age, $F(1, 52) = 5.35$, $p = .025$, $\eta^2 = .09$, indicating that 4-year-olds showed better recall than 3-year-olds (36% and 24%, respectively). There was also a trend toward a 3-way interaction involving sentence type, condition, and age group $F(1, 52) = 2.82$, $p = .099$, $\eta^2 = .05$. Post-hoc analyses showed that both 3- and 4-year-olds recalled generics more accurately than the quantifiers "all" and "most", $p = .065$ at age 3; $p < .01$ at age 4. However, for the generic/some condition, there was a strong age effect: 3-year-olds recalled generics better than "some" (38% vs. 10%), $p < .02$, whereas 4-year-olds recalled generics and "some" at equal rates (43% vs. 40%), $p > .70$.

We next looked at the absolute percentage of trials with opposite errors, again reporting only those effects involving age. We obtained a main effect of age group, $F(1, 52) = 5.58$, $p = .022$, $\eta^2 = .10$, indicating that 4-year-olds made more opposite errors than 3-year-olds ($Ms = 30\%$ and 20%, respectively). We also performed planned t-tests to determine if the asymmetry in recall (remembering quantified as generic more than the reverse) held up within each age group and wording condition separately. Results indicated that the asymmetry held up for "all" and "most" at both age groups, $ps < .02$, but that only 3-year-olds showed the asymmetry for "some" ($p = .078$); 4-year-olds did not show the asymmetry for "some" ($p > .16$).

We next examined the relative frequency of correct vs. opposite responses. For "all" and "most", both age groups show the same asymmetry between generics and quantifiers: for generics, correct responses exceed opposite errors, $ps < .05$, and for quantifiers, opposite errors exceed correct responses, $ps < .05$. For "some", however, we see a developmental pattern: whereas 3-year-olds also show the asymmetry (for generics, correct responses tended to exceed opposite errors, $p < .05$, one-tailed, and for "some", opposite errors tended to exceed correct responses, $p < .05$, one-tailed), 4-year-olds show a different pattern (for generics, correct responses show a non-significant tendency to exceed opposite errors, $p = .065$, one-tailed, but for "some", opposite errors are equal to correct responses, $p = 1.0$).

Finally, we examined opposite errors as a percentage of errors, again reporting only those effects involving age. There was a main effect of age group, $F(1, 45) = 5.74$, $p = .021$, $\eta^2 = .11$, indicating that 4-year-olds made more opposite errors than 3-year-olds. We also examined whether the generic/quantifier difference held up within each age group. For both 3-year-olds and 4-year-olds, quantifiers were more often recalled as generic than vice versa, $ps < .05$.

4.3. Discussion

As predicted, preschoolers tend to recall quantified statements as generics. Their recall for generics was much better than their recall for quantifiers in both age groups, and quantified statements were more likely to be recalled as generics than vice versa, even controlling for the overall error rates. Preschoolers had difficulty correctly recalling quantified statements: they correctly recalled them on only 18% of trials, whereas generics were correctly recalled on 41% of trials. These findings provide further support for the generics-as-default hypothesis.

The post-test data indicate that preschoolers have difficulty interpreting "most", even when it applies to a specific and limited set of items (i.e., crayons in a box). This replicates findings by Papafragou and Schwarz (2005/2006) and Barner et al. (2009) (though Halberda et al. (2008) suggests that preschoolers may have some limited understanding of "most"). Clearly, the preschool-aged participants in our study had difficulty interpreting "most" in the post-test, so our findings concerning recall for "most" with this age group should be interpreted cautiously. Importantly, however, we obtain the same pattern of results on the memory task with "all", where preschoolers display excellent performance on the post-test. Thus, children's tendency to recall quantified statements as generics, like that of adults, cannot be due to difficulty with the quantifier in question.

As we predicted, we obtained developmental differences in children's recall of "some" statements. 4-year-olds, like adults, were more successful in differentiating "some" statements from generics than they were in differentiating "all" and "most" statements from generics. 3-year-olds, however, tended to recall "some" statements as generics, just as they recalled "all" and "most" statements as generics. These results mirror Hollander et al.'s (2002) finding that 3-year-olds, but not 4-year-olds or adults, interpret category-wide "some" statements as generics.

Although more research is needed to establish the basis of this developmental difference, it is interesting to note that our post-test data also show a developmental difference between 3- and 4-year-olds in their comprehension of "some". Sixty percent of 4-year-olds in our study showed a scalar implicature effect – that is, if all the crayons were in the box, they denied that *some* crayons were in the box. In contrast, only 9% of 3-year-olds showed a scalar implicature effect. This suggests that 4-year-olds understand "some" as expressing a more bounded generalization, whereas for 3-year-olds, "some" may express a broader and less limited sort of generalization. Thus, 3-year-olds may interpret "some" as more clearly expressing a generalization about the kind at large, whereas 4-year-olds may interpret "some" as expressing a more limited claim about a few particular members of the kind. To the extent that "some" statements are understood to make claims about the kind in general, they may be more likely to be interpreted as generics.

5. Experiment 2B

Experiment 2A successfully replicated Experiment 1A using preschool aged participants. We thus sought to replicate Experiment 1B with similarly aged children. As with adults in 1B, we wondered whether preschoolers would recall quantified statements as generics even if they do not hear any generics in the learning phase. Since performance in the post-test was poor for "most", we paired "all" with "some" for this task.

Although we were primarily interested to see whether children would recall generics even if they do not hear them initially, this study also allowed us to check whether the results with "all" in 2A were due to some general tendency to confuse "all" statements with other kinds of statements. That is, we wished to check that the high rate of opposite errors with "all" in Experiment 2A was not simply due to a tendency to recall "all" as whatever other statement it is paired with. If "all" statements just get confused with whatever other statement they are paired with, then children should be as likely to recall "all" as "some" in this task as they were to recall "all" statements as generics in 2A. Comparing the rates of opposite errors across these studies would allow us to address this possible explanation. We were also interested to see whether 4-year-olds would show any tendency here to recall "some" statements as generics. Although they did not appear to do so in Experiment 2A, they might, as did adults in 1B, display a subtle tendency to do so in this study.

5.1. Method

5.1.1. Participants

Sixteen preschoolers (8 girls, 8 boys; mean age = 4.44 years; range = 3.42–5.06 years) were recruited from the same pool as in Experiment 2A. None of them participated in 2A. An additional 5 children (two 4-year-olds and three 3-year-olds) were tested but were dropped from analysis; 4 of them because they didn't remember at least one predicate from each wording condition, and 1 because she provided all her answers in Spanish.

5.1.2. Materials

The same materials as in Experiment 1B were used.

5.1.3. Procedure

The same procedure as in Experiment 2B was used, except that the quantifiers "all" and "some" were used (instead of "most" and "some", because "most" is difficult for young children to understand). The children received the same post-test as in Experiment 2A.

5.1.4. Coding

The same coding practice as in the previous studies was used. Agreement on the different coding categories ranged from 93% to 99%, and Cohen's Kappas ranged from .82 to .98.

5.2. Results

5.2.1. Post-test

Children performed extremely well on "all", correctly denying that 2/6 or 4/6 are all (both 92%), and correctly reporting that 6/6 is all (100%). Children also performed well on "some", when shown 2 or 4 out of 6, correctly reporting that 2/6 is some (75%) and that 4/6 is some (92%). Children showed mixed performance on "some" when shown 6/6, with 7 out of 15 children responding that 6/6 is some (47%). (One child had missing data on the post-test.) There was no relationship between age and performance on this post-test, though this may have been due to the small samples and restricted age range.

5.2.2. Recall of quantified statements

We conducted a series of paired t-tests to determine whether there were significant differences in recall for sentences presented with 'all' vs. 'some'. On the percentage of trials recalled correctly, children performed better with "some" (42%) than with "all" (16%), t-paired(15) = 2.89, p = .011. The percentage of trials recalled in the opposite form was higher for "all" (25%) than for "some" (2%), t-paired(15) = 3.31, p < .01. We also examined the relative frequency of correct vs. opposite errors. In the "all" condition, correct responses and opposite errors occurred at equal rates, t-paired(15) = −0.96, p > .3. Furthermore, in the "some" condition, correct responses were more frequent than opposite errors, t-paired(15) = 4.57, p < .001. Finally, the percentage of error trials recalled in the opposite form showed a trend to be higher for "all" (27%) than for "some" (9.5%), t-paired(13) = 1.92, p = .077.

We also conducted post-hoc tests on accuracy and percentage of opposite errors, to determine if the effects reported above held true for both younger and older preschoolers, by splitting the participants into two equal groups, a younger group (3.42–4.53, mean age 3.96) and an older group (4.61–5.06, mean age 4.87). (No analyses were conducted on the percentage of errors that were recalled as opposite, as the N of 14 was insufficient to split the group in two for that measure.) We predicted, based on the greater evidence for scalar implicatures among the older children, that they would be more likely to distinguish "some" from "all" than the younger children. This was indeed the case. Each of the effects reported above was significant among the older group (all ps < .01) but non-significant among the younger group (all ps > .2).

In order to see whether results concerning "all" in Experiment 2A might be due to "all" just being easily confused with whatever it is paired with, we compared the rates of opposite errors with "all" in Experiment 2A to the rates of opposite errors with "all" in this study. Rates of opposite errors with "all" tended to be lower in this study: 40% of trials were opposite errors in Experiment 2A but only 25% of trials in this study, $t(30)$ = 1.60, p = .06, one-tailed. This comparison suggests that the rate at which "all" was recalled as generic in Experiment 2A is not likely to be explained by a general tendency to confuse "all" with other statements. The rate of opposite recall for "some" was also lower here: 37% in Experiment 2A vs. only 2% in this study, $t(38)$ = 4.11, p < .001. This result further suggests that young children's errors in recalling quantifiers in Experiment 2A were largely due to a tendency to recall quantified statements as generics, rather than a base rate of erroneous recall.

The other primary purpose of this experiment was to examine whether quantified sentences are ever recalled as generic, even in the strong case when no generics are supplied during the learning phase. In contrast to Experiment 2A, where the only forms produced for "all" and "some" trials were those that had been presented in the original sentence block (i.e., the original quantifier, or else generic), here we found that participants consistently recalled quantified forms as generic. Over 60% of participants recalled a quantified sentence as generic at least once (N = 10 out of 16 participants). Notably, errors in which the "all" sentences were recalled as generic were equivalent to errors in which they were recalled in the opposite quantified form ("some"), p > .6, and errors in which the "some" sentences were recalled as generic were *higher* than errors in which they were recalled in the opposite quantified form ("all"), p < .05. Although this latter pattern was not significant when

examined within each age subgroup individually, due to reduced statistical power with the smaller *N*s, the same pattern obtained for both the younger and older subgroup.

5.3. Discussion

As found with the adult participants in Experiment 1B, preschoolers sometimes recalled statements as generics, despite not hearing a single generic in the learning phase. More specifically, they recalled "all" statements as generics as often as they recalled "all" as "some", and they recalled "some" statements as generics *more* often than they recalled "some" as "all". Even though every statement heard in the learning phase was quantified, participants were at least as likely to recall them as generics as they were to confuse one of the quantifiers with the other. This reflects a marked tendency to recall quantified statements as generics.

This experiment also confirmed that the high rates of recalling "all" as generic in Experiment 2A were not due to the quantifier "all" simply being easily confused with whatever it is paired with. The rates of opposite errors in this study for "all" tended to be lower than when "all" was paired with generics in Experiment 2A, suggesting again a tendency to recall "all" statements specifically as generics, rather than to just confuse them with other statements more generally.

Further, while "some" was remembered more accurately than "all" in both this study and in Experiment 2A, the rates of opposite recall for "some" were again lower in this study than in the latter experiment. As with "all", this finding suggests that "some" is more easily recalled as generic than confused with another quantifier, as is further supported by the fact that preschoolers were more likely to recall "some" as generic than as "all" in this study, despite never hearing a generic in the learning phase. These patterns held for 4-year-olds as well as 3-year-olds, suggesting that 4-year-olds, like adults, have a mild tendency to recall "some" statements as generics.

6. Experiment 3

Experiments 2A and 2B find that 3- and 4-year-olds, like adults, recall quantified statements as generics. These findings would seem to support the generics-as-default hypothesis. However, there is an alternative explanation available at least in the case of 3-year-olds – namely that they are simply forgetting the first word (i.e., the quantifier). Since 4-year-olds, like adults, tended to successfully recall "some" statements, it is unlikely that they were employing a 'dumb' strategy of simply dropping the first word. Such a strategy would apply equally well to "some" statements; thus, one would expect that "some" statements would also be recalled as generics, if the results were simply driven by this strategy. However, because 3-year-olds *did* recall "some" statements as generics, the possibility remains that they were simply forgetting the first word they heard. Experiment 3 was designed to address this possibility. In Experiment 3, 3-year-olds were presented with (positive) generics as before, but this time they were paired with the negative quantifier "no". We reasoned that if the 3-year-olds were simply forgetting the first word, then they should recall "no" statements as positive generics: for example, "no spiders have nine legs" would become "spiders have nine legs".[2] Previous work suggests that this experiment would be a particularly strong test, since at least adults show a bias towards recalling negative statements as affirmative statements, but not vice versa (Cornish & Wason, 1970; Fillenbaum, 1966; Mehler, 1963; but see Hörmann, 1971). Thus, if 3-year-olds did not recall "no" statements as positive generics, this would constitute clear evidence that they were not simply dropping the first word.

In contrast, if 3-year-olds were recalling "all", "most" and "some" statements as generics because of a cognitive bias towards interpreting quantified statements as generics, then "no" statements should either be correctly recalled as "no" statements, or else they should be recalled as *negative* generics (e.g., "spiders don't have nine legs"). If "no" statements were correctly recalled as "no" statements, then this would suggest that the previous findings did not just reflect a 'dumb' strategy of

---

[2] We are indebted to Adam Bowen for suggesting this way of testing whether 3-year-olds are simply forgetting the first word.

dropping the first word, but that the effect is limited to positive quantifiers such as "all", "most", and "some", and does not extend to negative quantifiers like "no."

However, if 3-year-olds recalled "no" statements as *negative* generics – that is, if "no spiders have nine legs" is recalled as "spiders don't have nine legs" – then this would suggest that the effect is not just limited to positive quantifiers. Since recalling such a negative generic involves restructuring the sentence to include a negation, it could not be due to a 'dumb' strategy of dropping the first word. Rather, it would reflect that even negative quantifiers such as "no" may be recalled as generics. Previous empirical (e.g., Leslie et al., 2011) and theoretical work (e.g., Leslie, 2007, 2008) only considered generics as defaults for *positive* quantifiers, so this experiment also represents a potential extension of the thesis to negative quantifiers. (Or, to use more precise semantic terminology, previous work has only considered monotone increasing quantifiers, and this experiment represents an extension of the thesis to monotone decreasing ones (Barwise & Cooper, 1981))

6.1. Method

6.1.1. Participants

Sixteen 3-year-olds (8 girls, 8 boys; mean age = 3.56; range = 3.08–3.97 years) were recruited from the same pool as in Experiment 2. None had participated in any of the prior studies. An additional 4 children were tested but were dropped from analysis; 3 because they failed to remember at least one predicate for each wording condition, and 1 because she only produced barking sounds.

6.1.2. Materials

The same pictures were used as in the previous studies; however, new facts were used for the quantifier "no", so that the novel facts would remain true. Since the facts used previously were true when associated with positive quantifiers, they would become false when associated with negative ones, and thus new facts were needed. A number of positive facts were also replaced, so that the negative and positive facts would be similar (e.g., "hippos eat plants" vs. "no hippos eat meat"). The new materials are shown in Table 2.

6.1.3. Procedure

The same procedure as in Experiment 2B was used, except that half the items were presented as positive generics, and the other half were presented as "no" statements.

Table 2
Items used in Experiment 3.

Category	Generic form	"No" form
Bear	Bears can bark	No bears can sing
Shark	Sharks breathe underwater	No sharks breathe in air
Rabbit	Rabbits keep food in their tummies	No rabbits keep food in their cheeks
Snail	Snails breathe through their side	No snails breathe through their mouths
Hippo	Hippos eat plants	No hippos eat meat
Skunk	Skunks go out at night	No skunks sleep at night
Spider	Spiders have eight legs	No spiders have nine legs
Bee	Bees have five eyes	No bees have six eyes
Chimp	Chimps can run fast	No chimps can swim
Rhino	Rhinos have horns made out of hair	No rhinos have horns made out of bone
Butterfly	Butterflies taste with their feet	No butterflies taste with their mouths
Frog	Frogs sleep all winter long	No frogs stay awake all winter long
Camel	Camels store fat in their humps	No camels store water in their humps
Snake	Snakes have tongues	No snakes have ears
Seahorse	Seahorses come out of their dads' tummies	No seahorses come out of their moms' tummies
Cat	Cats sweat through their paws	No cats sweat through their tails

6.1.4. Coding

Responses were coded as Positive Generic, Negative Generic, No, Positive Quantifier, or Other. Statements were coded as Positive Generic if they were in generic form (bare plural or indefinite singular) and did not contain any negations. If a statement was in generic form but did contain a negation (e.g., "spiders don't have nine legs"), then it was coded as Negative Generic. Whether a statement counted as a generic or not was determined by its subject noun phrase (NP). Occasionally, children produced responses that contained the quantifier "no", but in object position, with the subject NP in generic form, e.g. "squirrels have no stinking tails". These were coded as Negative Generics, since the quantifier "no" was attached to the object NP "stinking tails". If "no" modified the subject NP, then the statement was coded as No. If the statement contained a negative adverb of quantification such as "never", then it was coded as "no" also, since such an adverb could modify the subject NP at logical form (Heim, 1982; Lewis, 1975). Responses were coded as Positive Quantifier if they contained a positive quantifier ("all", "most" "some") in the subject, or a positive adverb of quantification ("always", "mostly", "sometimes"). All other statements were coded as Other. Transcripts were coded by two coders, with agreement across coding categories ranging from 97% to 100%, and Cohen's Kappas ranging from .92 to 1.0.

6.2. Results

As in the previous studies, children were more successful at recalling generic statements than quantified statements (see Fig. 3). Specifically, children correctly recalled generic sentences (52%) at a higher rate than "no"-quantified sentences (21%), t-paired(15) = 2.58, p = .021. Of primary interest was the rate at which children recalled "no" statements as *positive* generics, because a high rate of doing so would suggest that 3-year-olds are simply dropping the first word of the sentence. This rarely happened, however: "no" statements were recalled as positive generics on only 8% of trials, which was equal to the rate at which the (positive) generics were recalled as "no" statements (also 8%). This pattern held up when controlling for the rate of producing errors: the percentage of errors that were opposite for generic sentences (i.e., recalled as "no" statements) was equivalent to the percentage of errors that were opposite for "no" statements (i.e., recalled as positive generic), 15% and 11%, respectively, p > .6.
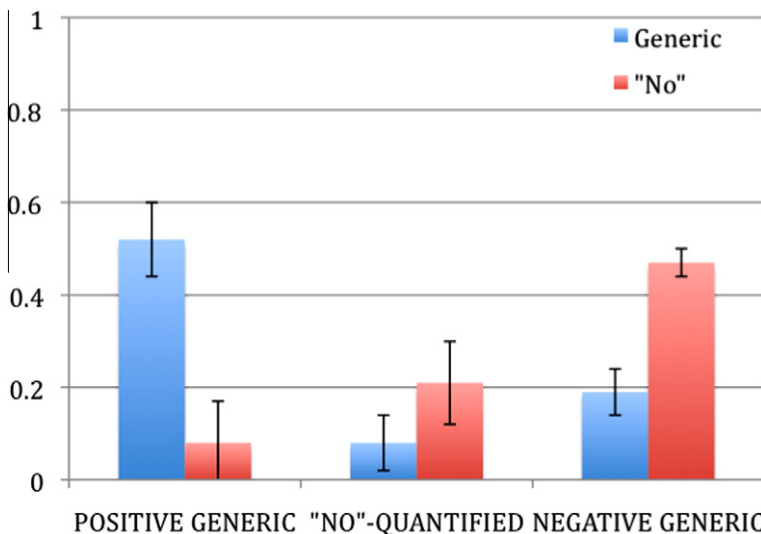
Fig. 3. Experiment 3, mean proportion of sentences recalled as positive generic (e.g., "Bears…"), "no"-quantified (e.g., "No bears…"), and negative generic (e.g., "Bears don't…").

Although "no" statements were rarely recalled as positive generics (8%), they were frequently recalled as *negative* generics (47% of trials), *t*-paired(15) = 3.93, *p* = .001. "No" statements were correctly recalled as "no" statements on 21% of trials, and there was a non-significant trend towards "no" statements being recalled as negative generics more often than they were correctly recalled as "no" statements, *t*-paired(15) = 1.66, *p* = .12. Further, 11 out of 16 participants recalled "no" statements as negative generics more often than they recalled "no" as "no," despite not hearing any negative generics in the learning phase. Only 4 participants showed the opposite pattern of recalling "no" statements correctly more often than recalling them as negative generics.

6.3. Discussion

"No"-quantified statements were rarely recalled as positive generics, which strongly suggests that 3-year-olds are not simply dropping the first word of the sentence in recall. If the previous results were due to such a strategy, then one would expect that "no" statements would be frequently recalled as positive generics. Instead, they were recalled as positive generics only to the same (low) extent that generics were recalled as "no" statements. These rates are naturally interpreted as the background rate at which the 3-year-olds genuinely misremembered negative statements as positive statements, and vice versa (which, interestingly, did not differ depending on whether the original statement was positive or negative, in contrast to, e.g., Cornish & Wason, 1970). Thus, the present results demonstrate that 3-year-olds were not simply forgetting the first word in the sentence, and thus that 3-year-olds' tendency to recall quantified statements as generics in Experiments 2A and 2B cannot be explained away by such a 'dumb' strategy.

Further, 3-year-olds have a tendency to recall "no" statements as negative generics, and they did so more than twice as often as they recalled "no" statements as "no" (i.e., in the form they had been originally presented, although this difference did not reach significance). Tellingly, 69% of participants more often recalled "no" statements as negative generics rather than as "no" statements, whereas only 25% showed the reverse tendency. This finding is notable because the participants did not hear *any* negative generics in the learning phase. This suggests that the tendency to recall quantified sentences as generic is not limited to positive (monotone increasing) quantifiers, but extends even to negative (monotone decreasing) quantifiers such as "no".

7. Experiment 4A

Experiments 4A and 4B were designed to test another alternative interpretation of the results. It is possible that participants interpret the quantified statements as quantifying not over *individual* members of the kind, but rather over *subkinds*. That is, perhaps participants – both preschoolers and adults – interpret, for example, "all bears climb trees" to mean *all kinds of bears climb trees* (e.g., grizzly bears climb trees, black bears climb trees, polar bears climb trees, and so on). If so, this would mean that participants were interpreting the quantified statements as making *generic* claims about all/most/some/none of the subkinds. If they interpret the quantified statements this way, then they should endorse statements with the quantifiers "all" or "most" as long as the predicate is generically true of all or most of the subkinds. Importantly, if this were the case, then the pattern of results presented in Studies 1–3 could be accounted for without reference to the hypothesis that generics are cognitively default relative to quantifiers.

It is unlikely that adults employed such interpretations. Leslie et al. (2011, Experiment 2B) solicited paraphrases of similar quantified statements, and found vanishingly little evidence of subkind-based interpretations. However, to our knowledge no previous work has investigated whether preschoolers may be inclined to interpret quantified statements in this way. Moreover, paraphrases may not provide the most sensitive test of adults' interpretations. Therefore, Experiment 4A was designed to directly test whether adults and preschoolers show any tendency to interpret universally quantified statements as pertaining to subkinds rather than individuals.

Participants (preschoolers and adults) were presented with item sets that each depicted eight members of a novel animal kind (e.g., gorps). Six of the eight instances had a distinctive feature

(e.g., spots) whereas the remaining two did not. Thus the stimuli were designed so that the generic statement (e.g., "Gorps have spots") would be likely deemed true, whereas the universal claim (interpreted as pertaining to the individual gorps; e.g., "All gorps have spots") would be deemed false. In one condition (the one-kind condition), the animals were all the same color and varied only in the distinctive feature, thus rendering a subkind based interpretation of the universal quantifier unlikely. Thus, participants should uniformly reject the universal statement. In the other condition (the subkind condition), however, the animals were divided into two groups of four, differing in color, with three out of four in each group having the relevant feature. Thus a subkind based interpretation of the universal quantifier would be available, if participants were inclined to employ it.

Crucially, if participants interpreted the universal as quantifying over subkinds, then they should agree that "All gorps have spots" in the subkind condition, because all subtypes of gorps have spots (even though not every individual does). Thus, if participants are inclined to give subkind interpretations to "all", they should endorse "all" statements more often in the condition where the subkinds are available than in the homogeneous condition where they are not. If, however, participants are not inclined to interpret universals as quantifying over subkinds, then they should treat the two conditions equivalently.

7.1. Method

7.1.1. Participants

Thirty-two preschoolers (16 girls, 16 boys; mean age = 4.42; range = 3.51–5.07) were recruited from the same pool as in Experiments 2 and 3. Four additional children were tested but dropped from analysis, due to a 'yes' answer bias (answering 'yes' to all questions, including warm-up questions). 42 adults (31 female, 11 male; mean age = 18.64; range = 18–21) were also recruited from the same pool as in Experiment 1. Sixteen preschoolers and 34 adults were assigned to the one-kind condition and 16 preschoolers and 8 adults were assigned to the subkind condition.

7.1.2. Materials

Ten item sets were employed, each consisting of drawings of 8 exemplars of a novel animal kind. In the one-kind condition, the 8 exemplars were identical, except for the target property. In the subkind condition, half the exemplars were one color, and the other half were another color, forming two subkinds; the two subkinds were spatially segregated. Four of the 10 sets were used in the warm-up task: for 2 warm-up sets, all 8 animals had the target property; for the other 2 warm-up sets, none of the 8 animals had the target property. The remaining six sets appeared in the test phase. In all of these sets, 6 out of 8 of the animals had the target property. In the subkind condition, the target property was equally distributed among members of the subkinds (i.e., 3 out of 4 animals in each subkind had the target property). See Fig. 4 for a sample stimulus set.

7.1.3. Procedure

Participants were each tested individually. They were shown the picture sets one at a time. For each set, the animals were introduced with a novel label (e.g., "These are gorps"). Participants were then asked to point to all instances (e.g., "Point to the gorps") and were corrected if they missed any. They were then asked a question about the target property, in either generic form ("Do gorps have humps?") or universally quantified form ("Do all gorps have humps?"). Question form was manipulated within-subjects, whereas the one-kind vs. subkind distinction was manipulated between-subjects.

The generic and "all" wording questions were blocked, and order of blocks was counterbalanced across participants. Thus, each block included two warm-up questions (one in which 0/8 instances depicted the target property; one in which 8/8 instances depicted the target property), followed by three test questions (in which 6/8 instances depicted the target property). Assignment of item set to wording block was counterbalanced across participants.
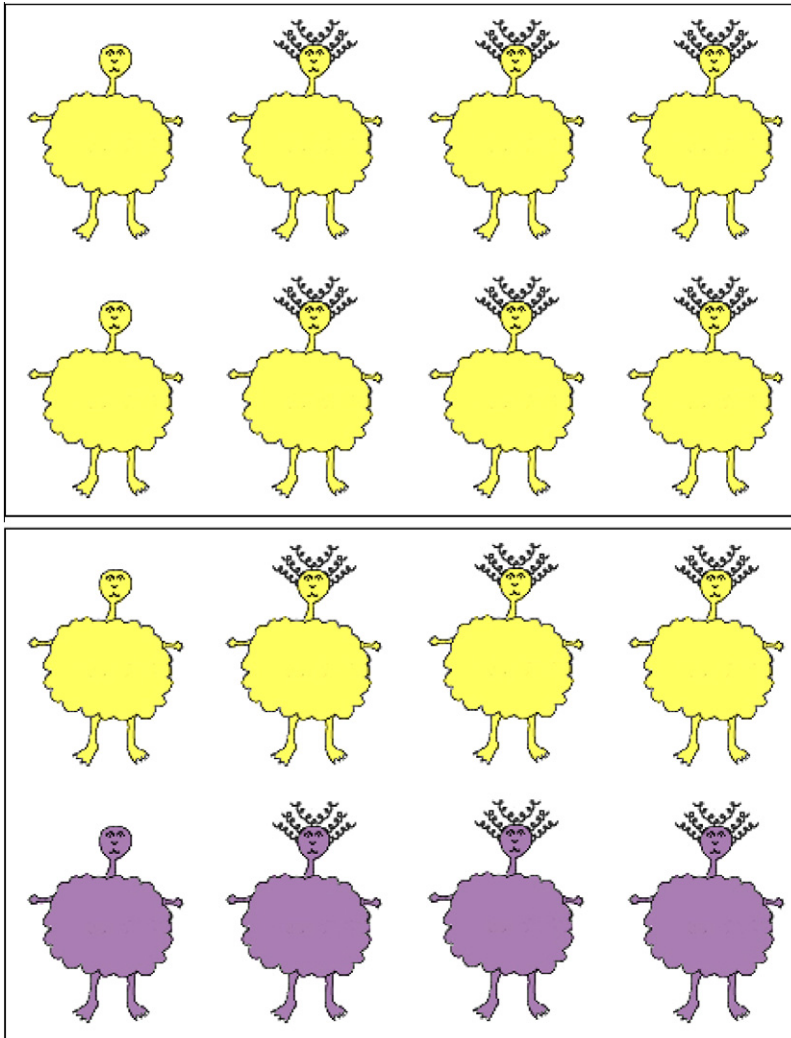
Fig. 4. Sample item sets, Experiment 4 (single subset and distinct subsets).

7.2. Results

Performance on the warm-up task was excellent across conditions and question wording (82% correct for preschoolers; 97% correct for adults), indicating basic understanding of the task. Of greater interest was the test phase. Responses to the test questions were entered into a three-way ANOVA, with wording (generic vs. "all") as a within-subjects factor, and age group and condition (one-kind vs. subkind) as between-subjects factors. There was a main effect of wording, $F(1,70) = 118.19$, $p < .001$, $\eta^2 = .63$, a main effect of age, $F(1,70) = 16.64$, $p < .001$, $\eta^2 = .19$. and a wording × age-group interaction, $F(1,70) = 46.31$, $p < .001$, $\eta^2 = .40$. Post-hoc tests reveal that at both age groups, participants said "yes" more often in response to the generic question than the "all" question, $ps < .005$, indicating that at no age was the quantifier "all" considered simply equivalent to the generic. Furthermore, children and adults did not significantly differ from one another in their responses to generic questions, endorsing them 78% and 86% of the trials, respectively. In contrast, there was a

sharp developmental difference in response to the "all" questions, with children endorsing them 58% of the trials and adults endorsing them <1% of the trials, *p* < .001.

Importantly, there were no significant effects involving condition (one-kind vs. subkind). If the results from Experiments 1–3 could be explained by a tendency to interpret universals as applying to subkinds, then the responses within at least one age group to universal questions should have differed by condition, that is, by whether a subkind interpretation was available or not. The results of this experiment do not provide any support for this hypothesis.

7.3. Discussion

If either children or adults showed a tendency to interpret "all" as quantifying over subkinds, they should have been more likely to respond "yes" to "all" questions in the subkind condition than in the one-kind condition. Experiment 4A did not find any evidence of this tendency. However, children did accept universals at a surprisingly high rate when only 6/8 exemplars had the target property. While this tendency was not found to differ between the one-kind and subkind conditions, it is in itself quite puzzling.

An interesting comparison point is children's very good performance on the 'all' comprehension post-tests in Experiments 2A and 2B, in which children rarely (less than 10% of trials) agreed that "all the crayons are in the box" when only 2/6 or 4/6 of the crayons were in the box. What might account for this difference? One explanation might be that there was something inherently puzzling in the novel materials used in Experiment 4A, or perhaps the difference between having only six crayons as opposed to eight animals was crucial to children's performance. The generics-as-default hypothesis, however, offers another possible explanation. Perhaps the difference in performance was due to the fact that children had to consider *kind-wide* universals in Experiment 4A ("all gorps"), whereas in the post-tests of Experiments 2A and 2B, the universal quantifier occurred in a partitive construction ("all *of the* crayons"), and thus only applied to a specific and limited set, namely the six crayons present in the available context. The generics-as-default hypothesis predicts that kind-wide quantified statements will sometimes be interpreted as generics; perhaps the results of Experiment 4A are due to children evaluating the *generic* instead of the kind-wide universal on approximately half the trials. That is, perhaps the results of Experiment 4A are examples of the same phenomenon documented by Hollander et al. (2002), Leslie et al. (2011), and Tardif et al. (2011). If so, this is an especially dramatic illustration, since the children had counter-examples right before them.

A full investigation of the phenomenon would constitute a more extensive set of experiments and so is left for future work. However, as an initial foray, we wished to test whether the high rate of accepting universals was due to the use of the kind-wide construction, or whether there was some feature of the stimuli used that was confusing to the children. To test this, we ran an adaptation of Experiment 4A using the universal partitive construction in place of the kind-wide universal.

8. Experiment 4B

8.1. Method

8.1.1. Participants

Sixteen preschoolers (8 girls, 8 boys; mean age = 4.39; range = 3.29–5.79) and 10 adults (5 female, 5 male; mean age = 19.1; range = 18–21) were recruited from the same subject pool as in Experiment 4A. An additional three children were tested, but were dropped from analysis due to a 'yes' bias in two cases, and a 'no' bias in the third (i.e., providing the same response [either 'yes' or 'no'] to every question, including the warm-up questions).

8.1.2. Materials

The same materials were used here as in the one-kind condition in Experiment 4A.

8.1.3. Procedure

The procedure was the same as the one-kind condition of Experiment 4A, with two differences: no generic questions were asked (because generics only apply to entire kinds, not to specific sets (Carlson & Pelletier, 1995)), and the universal questions involved partitive constructions (e.g., "Do all of these Gorps have humps?") so that they quantified over just the particular individuals depicted in the drawings.

8.2. Results

As in Experiment 4A, performance on the warm-up trials was excellent (80% correct among preschoolers; 95% correct among adults). More importantly, participants also performed extremely well on the test trials. Adults rejected the universal ("all of these") on 100% of trials in the test conditions. Crucially, children's responses to the test questions were significantly improved from Experiment 4A, with children incorrectly agreeing to the universal when only 6/8 have the target property on only 21% of trials, as compared with 58% of trials in Experiment 4A, $t(46) = 2.91$, $p < .01$.

8.3. Discussion

Children performed much better when the universal applied only to a specific set (e.g., "all of these gorps"), as opposed to when the universal applied kind-wide (e.g., "all gorps"). Children correctly rejected the universal on 79% of trials in Experiment 4B, which is comparable to the 80% of trials on which they answered correctly on the warm-up (e.g., agreeing that all the gorps have wings when 8 out of 8 of them do) – in contrast to 42% accuracy on "all" trials in Experiment 4A. This comparison suggests that the tendency to agree to the universal in 4A was not due to some confusing feature of the novel stimuli, but rather was due to "all" applying to the entire kind, rather than to only the particular depicted instances. More work would need to be done to be fully certain of the explanation of this finding. However, it should be noted that this is exactly the sort of finding that the generics-as-default hypothesis would predict. The generics-as-default hypothesis predicts that both children and adults will sometimes interpret kind-wide quantified statements as generics (Hollander et al., 2002; Leslie et al., 2011; Tardif et al., 2011). If the children in Experiment 4A were doing this, it would explain the high rate of agreement to the kind-wide universals, because the children were often instead evaluating the generic, which was itself robustly accepted.

Experiments 4A and 4B did not find any evidence that either children or adults are inclined to interpret quantified statements as applying to subkinds rather than individuals, which would have been an alternative explanation of the results of Experiments 1–3. Instead, Experiments 4A and 4B together turned up what is potentially a further piece of support for the generics-as-default hypothesis, which future work will need to explore in more detail.

9. General discussion

Several researchers in psychology and philosophy have recently proposed that generic sentences may articulate cognitively default generalizations (Csibra & Gergely, 2009; Gelman, 2010; Gelman & Brandone, 2010; Hollander, Gelman, & Raman, 2009; Leslie, 2007, 2008; Leslie et al., 2011). In contrast, quantified statements may express cognitively more sophisticated generalizations. The results presented here lend support to this generics-as-default hypothesis, as they indicate that both preschoolers and adults have a tendency to recall quantified statements as generics.

In addition to characterizing the results presented here, the generics-as-default hypothesis would explain why Hollander et al. (2002) and Tardif et al. (2011) found that English- and Mandarin speaking preschoolers seem to rely on their evaluation of the corresponding generic when asked to evaluate category-wide quantified statements – an effect that Experiment 4 here suggests may be quite powerful indeed, at least in the case of "all". It is also consistent with findings by Leslie et al. (2011), which suggest that adults may similarly rely on their judgments of generics to evaluate certain universally quantified statements. Leslie (2008) also argues that such a hypothesis would explain why no known

language has a word (e.g., "gen") that signals a generic statement in the way that a quantifier such as "all" signals a universal statement.

Certain adult reasoning errors would also be explicable on this hypothesis, for example Jönsson and Hampton's (2006) finding that adults judge "all ravens are black" to be more likely to be true than "all young jungle ravens are black". From a logical point of view, this is an error, because the young jungle ravens form a subset of the ravens, and so if "all ravens are black" is true, then "all young jungle ravens are black" must also be true. However, if adults were interpreting these universal statements as generics, then their judgments would be appropriate: "ravens are black" could be true even if young jungle ravens weren't black – just as "tigers are striped" is true even though albino tigers aren't striped (see Connolly, Fodor, Gleitman, & Gleitman, 2007; see also Cimpian, Gelman, et al., 2010). Thus Jönsson and Hampton's findings would be easily explained if category-wide universal statements tended to be interpreted as generics by adults. A similar explanation may also apply to Sloman's (1993, 1998) findings concerning inferences about categories.

9.1. Is it just that generics are shorter?

One possible alternative explanation for the findings reported here could be that people simply have a general tendency to recall sentences in shortened form. That is, perhaps generics are favored in recall because they are shorter by one word than quantified sentences. This explanation seems implausible for a number of reasons. First, it is not clear how such an explanation could explain why "all" and "most" but not "some" were recalled as generics by adults and 4-year-olds, because in all three cases the generic is shorter than the quantified statement. Further, Experiment 3 indicates that even 3-year-olds were not simply 'dropping the quantifier,' since they tended to recall "no" statements as *negative* generics, despite the fact that negative generics are not inherently shorter than "no" statements (as "no" is replaced by "don't" here).

Nonetheless, to further examine this question, we looked to see whether participants displayed a general tendency to shorten the lengths of the *predicates* in recall. If generic recall was simply due to an overall tendency to shorten, then one would expect that predicates would also tend to be shortened in recall. We did not find consistent evidence of this, however; in most cases, participants produced predicates of equal length to those used in the learning phase. Further, although participants in Experiments 1A and 2A were more likely to shorten rather than lengthen a predicate, participants in Experiments 1B and 2B did not show such a tendency. The breakdown from these four studies is as follows. Study 1A: Adults kept predicates the same length on 73% of trials, shortened predicates on 17% of trials, and lengthened on 10% of trials (shorter > longer, $p < .01$). Study 1B: Adults kept predicates the same length on 74% of trials, shortened predicates on 14% of trials, and lengthened on 12% (shorter = longer, $p > .4$). Study 2A: Children kept predicates the same length on 43% of trials, shortened predicates on 35% of trials, and lengthened on 22% (shorter > longer, $p < .05$). Study 2B: Children kept predicates the same length on 52% of trials, shortened predicates on 22% of trials, and lengthened on 26% (shorter = longer, $p > .5$). Thus we did not find any consistent tendency to shorten predicates in recall, as would be expected if our data were to be explained by a general tendency to shorten.

A further question is whether the tendency to recall a sentence as a generic is affected by its length. To examine this, we analyzed the likelihood of recalling a sentence as a generic as a function of its length (coded as short (2 or 3 words), medium (4 words), or long (5 or 6 words)). Again, this analysis did not reveal a consistent pattern: although adults were more likely to recall longer sentences as generics, preschoolers were not. (The rates of recall as generic was as follows: for Study 1A (adults), 44% for short sentences, 56% for medium sentences, and 60% for long sentences; for Study 1B (adults), 7% for short sentences, 11% for medium sentences, and 22% for long sentences; for Study 2A (preschoolers), 39% for short sentences, 41% for medium sentences, and 41% for long sentences; for Study 2B (preschoolers), 27% for short sentences, 25% for medium sentences, and 30% for long sentences.)

Again, the fact that there was not a consistent pattern of generic recall as a function of sentence length argues against the idea that our results simply depend on a tendency to shorten sentences in recall. Given that preschoolers' rate of generic recall was not affected by sentence length, this is unlikely to be the explanation of the results. It is interesting, however, that adults showed a somewhat greater tendency to recall longer sentences as generic. One possible explanation for the finding is con-

sistent with the generics-as-defaults hypothesis, namely, that this tendency may be due to longer sentences imposing somewhat higher processing demands. That is, if processing and remembering a longer sentence imposes higher cognitive demands, then this may lead to an increased tendency to default to the generic. Some indirect support for this possibility is found in Meyer, Gelman, and Stilwell (2011), who found that adults were more likely to evaluate universals as generics when time constraints were introduced. Further research is needed to address this question; in particular it will be important to examine whether increased cognitive load influences adults' tendency to recall quantified statements as generics.

9.2. Developmental trajectories in the recall of quantified statements

Adults showed a definite overall tendency to recall quantified statements as generics. However, in the case of "some" and to some extent "all", the findings presented here reflect an interesting developmental trajectory: younger children are *more* inclined than adults to recall these quantified statements as generics. This result is what the generics-as-default hypothesis would predict: the youngest children are the most susceptible to defaulting to the generic interpretation. This was evident for "all": adults were equally likely to recall "all" as generic as they were to recall it correctly, whereas preschoolers were more likely to recall "all" as generic.

In the case of "some", the developmental trajectory was quite dramatic, with 3-year-olds showing a strong tendency to recall "some" as generic, whereas 4-year-olds and adults showed very little tendency to do so. It is possible that this developmental change with "some" is connected to a developing appreciation for scalar implicature, meaning that 3-year-olds may interpret "some" as indicating more open-ended and potentially kind-wide generalizations than 4-year-olds and adults do. This is only a tentative interpretation, however, and further investigation into 3- and 4-year-olds' interpretations of "some" is needed before one could be confident of such an explanation. One way to test the hypothesis would be to examine whether individual performance with "some" on the memory task corresponds to appreciation of scalar implicature. In the current study, scalar implicature appreciation was assessed based on a single trial of the post-test, and so we did not have sufficient power to examine correlations at the individual level. However this is a question that will be important to assess in future work.

The fact that developmental changes occurred with recall of "all" and "some" suggests that there may also be a more general developmental trajectory toward increasingly suppressing the tendency to default to the generic in recall. In light of this, it is interesting to note how poorly even adults performed in recalling "most". A possible explanation of this may be that "most" is simply a more difficult quantifier to process than "all" and "some", as previous researchers have suggested (e.g., Barner et al., 2009; Barwise & Cooper, 1981; Papafragou & Schwarz, 2005/2006).

9.3. Conclusion

If the generics-as-default hypothesis is indeed correct, there are a number of implications for understanding human cognition. The way we most naturally generalize information from individuals to kinds is of central important in cognitive psychology (Prasada, 2000; Smith, 1989). The generics-as-default hypothesis claims that our most basic and automatic way of generalizing such information involves generic generalizations, rather than quantificational ones. Although the studies reported here did not focus on the exact nature of generic generalizations, a number of recent studies suggest that generic generalizations, unlike quantificational generalizations, do not depend solely on statistical information such as prevalence (Cimpian, Brandone, et al., 2010; Cimpian, Gelman, et al., 2010; Khemlani et al., 2007, 2009, in press; Leslie, 2007). Although quantified statements such as "most lorches have feathers" are accepted or rejected depending on the percentage of lorches that have feathers, adults' acceptance/rejection of generics such as "lorches have feathers" depends on more complex factors, such as whether the feathers are potentially dangerous (Cimpian, Brandone, et al., 2010; Leslie, 2007, 2008). Similarly, adults are more likely to endorse generics such as "dondrets have tails" if having a tail is a characteristic or essential property of dondrets that emerges with maturation, than if only infant dondrets have tails – *even* if there are more infant dondrets, and therefore more dondrets

with tails. Quantifiers such as "most" do not follow such patterns, but rather depend only on whether the majority of dondrets have tails (Cimpian, Gelman, et al., 2010).

If generic generalizations are indeed cognitively more basic than quantificational ones, then this would mean that our fundamental way of generalizing information is sensitive to more than just statistical distributions of properties. Rather, our basic general judgments would be sensitive to factors such as whether the property to be generalized is potentially dangerous (Cimpian, Brandone, et al., 2010; Leslie, 2007, 2008) or whether it is an essential property that emerges with development (Cimpian, Gelman, et al., 2010; Gelman, 2003). Such a hypothesis, if correct, would fit well with theory-based accounts of cognition (e.g., Carey, 2009; Gelman, 2003; Gopnik & Meltzoff, 1997; Gopnik & Wellman, 1994), according to which general knowledge about kinds is arranged according to causal structures, and is driven by more than simple statistical distributions of properties.

Acknowledgments

References

Barner, D., Chow, K., & Yang, S. (2009). Finding one's meaning: A test of the relation between quantifiers and integers in language development. *Cognitive Psychology, 58*, 195–219.
Barwise, J., & Cooper, R. (1981). Generalized quantifiers and natural language. *Linguistics and Philosophy, 4*, 159–219.
Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language, 51*, 437–457.
Bower, G. H., Black, J. B., & Turner, T. J. (1979). Scripts in memory for text. *Cognitive Psychology, 11*, 177–220.
Carey, S. (2009). *The origin of concepts*. New York: Oxford University Press.
Carlson, G. (1977). Reference to kinds in English. Ph.D. dissertation, University of Massachusetts, Amherst.
Carlson, G., & Pelletier, F. J. (1995). *The generic book*. Chicago: University of Chicago Press.
Chomsky, N. (2000). *New horizons in the study of language and mind*. Cambridge: Cambridge University Press.
Cimpian, A., Brandone, A. C., & Gelman, S. A. (2010). Generic statements require little evidence for acceptance but have powerful implications. *Cognitive Science, 34*, 1452–1482.
Cimpian, A., Gelman, S. A., & Brandone, A. C. (2010). Theory-based considerations influence the interpretation of generic sentences. *Language and Cognitive Processes, 25*, 261–276.
Connolly, A., Fodor, J., Gleitman, L. R., & Gleitman, H. (2007). Why stereotypes don't even make good defaults. *Cognition, 103*, 1–22.
Cornish, E. R., & Wason, P. C. (1970). The recall of affirmative and negative sentences in an incidental learning task. *Quarterly Journal of Experimental Psychology, 22*, 109–114.
Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences, 13*, 148–153.
Dahl, O. (1985). *Tense and aspect systems*. Oxford: Blackwell.
de Swart, H. (1993). *Adverbs of quantification: A generalized quantifier approach*. New York and London: Garland.
Ehrlich, M. F. (1975). Cumulative learning and long term retention of sentences. *Acta Psychologica, 39*, 241–250.
Fillenbaum, S. (1966). Memory for gist: Some relevant variables. *Language and Speech, 9*, 217–227.
Gelman, S. A. (2003). *The essential child*. New York: Oxford University Press.
Gelman, S. A. (2010). Generics as a window onto young children's concepts. In F. J. Pelletier (Ed.). *Kinds, things, and stuff: The cognitive side of generics and mass terms. New directions in cognitive science* (Vol. 12). New York: Oxford University Press.
Gelman, S. A., & Brandone, A. C. (2010). Fast-mapping placeholders: Using words to talk about kinds. *Language Learning and Development, 6*, 223–240.
Gelman, S. A., Goetz, P. J., Sarnecka, B. S., & Flukes, J. (2008). Generic language in parent–child conversations. *Language Learning and Development, 4*, 1–31.
Gelman, S. A., Hollander, M., Star, J., & Heyman, G. D. (2000). The role of language in the construction of kinds. In D. Medin (Ed.). *Psychology of learning and motivation* (Vol. 39, pp. 201–263). New York: Academic Press.
Gelman, S. A., & Raman, L. (2003). Preschool children use linguistic form class and pragmatic cues to interpret generics. *Child Development, 74*(1), 308–325.
Gelman, S. A., & Raman, L. (2007). This cat has nine lives? Children's memory for genericity in language. *Developmental Psychology, 43*(5), 1256–1268.
Gopnik, A., & Meltzoff, A. N. (1997). *Words, thoughts, and theories*. Cambridge, MA: Bradford Books/MIT Press.
Gopnik, A., & Wellman, H. (1994). The theory theory. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture*. NY: Cambridge University Press.

Graham, S. A., Nayer, S. L., & Gelman, S. A. (2011). Two-year-olds use the generic/non-generic distinction to guide their inferences about novel kinds. *Child Development, 82*, 493–507.

Halberda, J., Taing, L., & Lidz, J. (2008). The development of "most" comprehension and its potential dependence on counting-ability in preschoolers. *Language Learning and Development, 4*, 99–121.

Heim, I. (1982). The semantics of definite and indefinite noun phrases. Ph.D. dissertation, University of Massachusetts at Amherst.

Holyoak, K. J., & Morrison, R. G. (Eds.). (2005). *The Cambridge handbook of thinking and reasoning*. Cambridge, UK: Cambridge University Press.

Hollander, M. A., Gelman, S. A., & Raman, L. (2009). Generic language and judgments about category membership: Can generics highlight properties as central? *Language and Cognitive Processes, 24*, 481–505.

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

Hörmann, H. (1971). Semantic factors in negation. *Psychologische Forschung, 35*, 1–16.

Huang, Y., & Snedeker, J. (2009). Semantic meaning and pragmatic interpretation in 5-year-olds: Evidence from real-time spoken language comprehension. *Developmental Psychology, 45*, 1723–1739.

Hurewitz, F., Papafragou, A., Gleitman, L., & Gelman, R. (2006). Asymmetries in the acquisition of numbers and quantifiers. *Language Learning and Development, 2*, 77–96.

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General, 110*, 306–340.

Jönsson, M. L., & Hampton, J. A. (2006). The inverse conjunction fallacy. *Journal of Memory and Language, 55*, 317–334.

Keil, F. C. (1989). *Concepts, kinds, and cognitive development*. Cambridge, MA: MIT Press.

Khemlani, S., Leslie, S. J., Glucksberg, S., & Rubio-Fernandez, P. (2007). Do ducks lay eggs? How humans interpret generic assertions. In *Proceedings of the 29th annual conference of the cognitive science society*. Nashville, TN: Cognitive Science Society.

Khemlani, S., Leslie, S. J., & Glucksberg, S. (2009). Generics, prevalence, and default inferences. In *Proceedings of the 31st annual conference of the cognitive science society*. Austin, TX: Cognitive Science Society.

Khemlani, S., Leslie, S. J., & Glucksberg, S. (in press). Inferences about members of kinds: The generics hypothesis. *Language and Cognitive Processes*.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction–integration model. *Psychological Review, 95*, 163–182.

Kratzer, A. (1995). Stage-level and individual-level predicates. In G. Carlson & J. Pelletier (Eds.), *The generic book* (pp. 125–175). Chicago: Chicago University Press.

Krifka, M., Pelletier, F. J., Carlson, G., ter Meulen, A., Chierchia, G., & Link, G. (1995). Genericity: An introduction. In G. Carlson & F. J. Pelletier (Eds.), *The generic book* (pp. 1–125). Chicago: Chicago University Press.

Leslie, S. J. (2007). Generics and the structure of the mind. *Philosophical Perspectives, 21*, 375–405.

Leslie, S. J. (2008). Generics: Cognition and acquisition. *The Philosophical Review, 117*, 1–49.

Leslie, S. J., Khemlani, S., & Glucksberg, S. (2011). All ducks lay eggs? The generic overgeneralization effect. *Journal of Memory and Language, 65*, 15–31.

Lewis, D. (1975). Adverbs of quantification. In *Formal semantics of natural language* (pp. 3–15). Cambridge: Cambridge University Press.

Macrae, C. N., Hewstone, M., & Griffiths, R. J. (1993). Processing load and memory for stereotype-based information. *European Journal of Social Psychology, 23*, 77–87.

Margolis, E., & Laurence, S. (1999). *Concepts: Core readings*. Cambridge, MA: MIT Press.

McCarthy, R. A. (1995). *Semantic knowledge and semantic representations*. Hove, UK: Taylor & Francis.

Mehler, J. (1963). Some effects of grammatical transformations on the recall of English sentences. *Journal of Verbal Learning and Verbal Behavior, 2*, 346–356.

Meyer, D. E. (1970). On the representation and retrieval of stored semantic information. *Cognitive Psychology, 1*, 242–299.

Meyer, M., Gelman, S. A., & Stilwell, S. M. (2011). Generics are a cognitive default: Evidence from sentence processing. In *Proceedings of the 33rd annual conference of the cognitive science society*. Boston, MA: Cognitive Science Society.

Murphy, G. L. (2002). *The big book of concepts*. Cambridge, MA: MIT Press.

Noveck, I. (2001). When children are more logical than adults: Experimental investigations of scalar implicatures. *Cognition, 78*, 165–188.

Papafragou, A., & Schwarz, N. (2005/2006). Most wanted. *Language Acquisition, 13*, 207–251.

Papafragou, A., & Musolino, J. (2003). Scalar implicatures: Experiments at the semantics–pragmatics interface. *Cognition, 86*, 253–282.

Pouscoulous, N., Noveck, I., Politzer, G., & Bastide, A. (2007). Processing costs and implicature development. *Language Acquisition, 14*, 347–375.

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

Prasada, S., & Dillingham, E. (2006). Principled and statistical connections in common sense conception. *Cognition, 99*, 73–112.

Prasada, S., & Dillingham, E. (2009). Representation of principled connections: A window onto the formal aspect of common sense conception. *Cognitive Science, 33*, 401–448.

Rosch, E. H. (1973). Natural categories. *Cognitive Psychology, 4*, 328–350.

Sloman, S. A. (1993). Feature-based induction. *Cognitive Psychology, 25*, 231–280.

Sloman, S. A. (1998). Categorical inference is not a tree: The myth of inheritance hierarchies. *Cognitive Psychology, 35*, 1–33.

Smith, E. E. (1989). Concepts and induction. In M. I. Posner (Ed.), *Foundations of cognitive science* (pp. 501–526). Cambridge, MA: MIT Press.

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts*. Cambridge, MA: Harvard University Press.

Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review, 1*, 214–241.

Tabachnick, B. G., & Fidell, L. S. (1989). *Using multivariate statistics*. Philadelphia, PA: Harper & Row.

Tardif, T., Gelman, S. A., Fu, X., & Zhu, L. (2011). Acquisition of generic noun phrases in Chinese: Learning about lions without an "-s". *Journal of Child Language, 30*, 1–32.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 381–402). New York: Academic Press.