Research Article

Psychological Science
2021, Vol. 32(2) 189–203

DOI: 10.1177/0956797620953132
www.psychologicalscience.org/PS

The Unintended Consequences of the Things We Say: What Generic Statements Communicate to Children About Unmentioned Categories

TC

Kelsey Moty and Marjorie Rhodes
Department of Psychology, New York University

Abstract

Adults frequently use generic language (e.g., “Boys play sports”) to communicate information about social groups to children. Whereas previous research speaks to how children often interpret information about the groups described by generic statements, less is known about what generic claims may implicitly communicate about unmentioned groups (e.g., the possibility that “Boys play sports” implies that girls do not). Study 1 (287 four- to six-year-olds, 56 adults) and Study 2 (84 four- to six-year-olds) found that children as young as 4.5 years draw inferences about unmentioned categories from generic claims (but not matched specific statements)—and that the tendency to make these inferences strengthens with age. Study 3 (181 four- to seven-year-olds, 65 adults) provides evidence that pragmatic reasoning serves as a mechanism underlying these inferences. We conclude by discussing the role that generic language may play in inadvertently communicating social stereotypes to young children.

Keywords

generic statements, pragmatics, conceptual development, social groups, open data, open materials, preregistered

Received 8/29/19; Revision accepted 7/14/20

A child told that “girls wear pink,” “a boy doesn’t cry,” or “scientists discover new things” can learn (or mislearn) much about the world. These statements—in contrast to other forms of phrasing (e.g., “Some girls wear pink,” “Peter never cries,” “This scientist discovered something new”)—communicate generic information about what members of these groups are usually like, what members of these groups are supposed to be like, or what makes them as a group unique in comparison with other groups (Carlson, 1977; Carlson & Pelletier, 1995). By the age of 4 years, and in some cases earlier, children assume from statements such as “Boys love to wrestle” that *boys* is an informative way of grouping people (Rhodes, Leslie, Bianchi, & Chalik, 2018), that many boys love to wrestle (e.g., Brandone, Gelman, & Hedglen, 2015; Gelman, Ware, & Kleinberg, 2010; Rhodes, Leslie, & Tworek, 2012), and that they do so because of something intrinsic to being a boy (Cimpian & Erickson, 2012; Cimpian & Markman, 2009, 2011; Gelman et al., 2010; Rhodes et al., 2012). These assumptions are often wrong—adults use generic language to describe rare properties, too (as in, “Mosquitoes carry West Nile virus,” which is true of less than 1% of mosquitoes; Cimpian, Brandone, & Gelman, 2010; Cox, 2004), as well as to refer to extrinsic causal relations (as in, “Girls wear pink,” which is the result of marketing and not anything intrinsic to girls or the color itself; Cimpian & Salomon, 2014).

Here, we consider the possibility that children’s (sometimes inaccurate) interpretations of generic claims extend even further—to shape their beliefs about groups not mentioned at all. That is, when hearing, “Girls are good at drawing,” children may assume that a new girl will be a good drawer. But will they also then assume that a new boy will be bad? If so, these generic claims may inadvertently communicate social

Corresponding Author:
Kelsey Moty, New York University, Department of Psychology
E-mail: moty@nyu.edu

stereotypes to young children, a possibility important to consider given the frequency of generic claims in everyday parent–child conversations (Gelman, Goetz, Sarnecka, & Flukes, 2008; Graham, Nayer, & Gelman, 2011) and the ubiquity of generic claims across languages (Leslie, 2008).

Whether generic claims communicate social stereotypes to young children might depend on children's beliefs about why a speaker chose these words (Clark, 1996; Grice 1975; Horn, 1984). For example, if children (a) believe that the speaker knows a lot about both boys and girls and, therefore, (b) could have knowledgeably made claims about both, then children could interpret the speaker's decision to say that "Girls are good at drawing" as them intending to imply that boys are not (Jaswal & Neely, 2006; Koenig & Harris, 2005; Shafto, Goodman, & Frank, 2012). This is particularly the case when the speaker could as easily have said something else, such as "Boys and girls are good at drawing" or "Kids are good at drawing." Similarly, if the speaker saw a particular girl doing a great drawing but chose to make a general claim about girls (e.g., saying "Girls are good . . ." rather than "This girl is good . . ."), then this decision could be interpreted as intent to imply that the particular girl's behavior is an example of more general trends.

This type of reasoning—referred to as *pragmatic reasoning* or *pragmatic inferencing*—can be difficult for young children. For instance, children before the age of 7 struggle to realize that when someone says, "Jim ate some of the cookies," this means that he likely did not eat all of them (Huang & Snedeker, 2009; for examples with other quantifiers, see Noveck, 2001; Papafragou & Musolino, 2003). Yet this appears to be in part because children do not quite understand the meanings of "some" and "all" and fail to consider that *all* was an alternative word that the speaker could have chosen (Barner, Brooks, & Bale, 2011). In simplified contexts that either (a) do not require knowing the meaning of these quantifiers or (b) make clearer the alternatives that speakers could have said, preschool-age children engage in sophisticated pragmatic reasoning (e.g., Jara-Ettinger, Floyd, Huey, Tenenbaum, & Schulz, 2020; Stiller, Goodman, & Frank, 2015). In contrast to quantified claims (using *all* or *some*), generic claims are understood by young children, who can distinguish between generic claims and more specific ones by the age of 2.5 years (e.g., Graham, Gelman, & Clarke, 2016; Graham et al., 2011). Therefore, we hypothesized that children might draw sophisticated pragmatic inferences about the intended meanings of generic claims by quite early in childhood.

The aims of the present work were twofold: to identify (a) what children infer, if anything, about unmentioned groups from generic claims (Studies 1 and 2) and (b) whether children's inferences about unmentioned groups rely on their assumptions about a speaker's mental states (Study 3). To examine these questions, we used a novel social dichotomy between "zarpies" and "gorps" (marked by clothing color) so that participants would not be biased by preexisting stereotypes. Doing so provided a strong test of what generic claims communicate about unmentioned groups. We examined children between the ages of 4 and 6 years because during this time, children's pragmatic reasoning develops (e.g., Huang & Snedeker, 2009; Stiller et al., 2015), generic claims about social groups are commonly said to children (e.g., Gelman, Taylor, & Nguyen, 2004; Rhodes & Leslie, 2018), and children's representations of social categories undergo many changes (Liberman, Woodward, & Kinzler, 2017).

Study 1

This study examined whether participants expect properties described as true of one group (using generic or specific language) to extend to other individuals from that group and to an unmentioned group.

Method

Preregistered hypotheses. Following previous research, we hypothesized that children should more readily assume after hearing generic claims such as "Zarpies are good at baking pizzas" that other zarpies are good at baking pizzas. Critically, if generic claims also communicate information about unmentioned groups (e.g., gorps), then children should respond that gorps are not good at baking pizzas. In contrast, children should be less sure about the status of gorps after hearing the specific claim, "This zarpie is good at baking pizzas." We planned to examine age-related trends in children's inferences about both groups to distinguish among three possibilities: (a) These inferences emerge early and robustly with no changes across age, (b) these inferences emerge at a young age but become more robust across early childhood, or (c) these inferences emerge later in development. An additional adult sample was tested for comparison. The study design and analytic code to test these hypotheses were preregistered and can be found on OSF at https://osf.io/c3fpv.

Participants. The sample consisted of 56 adults (37% women, M = 38.1 years) and 287 children (57% girls) between the ages of 4 and 6 years (112 four-year-olds, M = 4.59 years; 83 five-year-olds, M = 5.40 years; 92 six-year-olds, M = 6.44 years; note that we examined children in binned age groups for analyses that also included adults but continuously in analyses that examined only children). Of those parents who provided racial and ethnic demographic information for their children (n = 248), 56% of them identified their child as White, 15% as Asian, 13% as Black/African American, 2% as Middle Eastern or North African, and 15% as mixed or biracial; 22% of these participants were also Hispanic/Latino. We recruited children from local public preschools, elementary schools, and a children's museum in New York City. Adult participants were recruited via Amazon's Mechanical Turk. Adult participants predominantly identified as White (87%; other racial groups present in our sample each composed less than 5% of the sample); 10% of adult participants identified as Hispanic/Latino. Approximately half of participants were assigned to one of two language conditions (generic language vs. specific language).

We preregistered a sample size of 60 participants per age bin (4-year-olds, 5-year-olds, 6-year-olds, and adults) with half in each language condition (planned total N = 240 [180 children, 60 adults]) on the basis of the sample sizes used in comparable studies exploring pragmatic abilities in early childhood (e.g., Horowitz & Frank, 2016; Stiller et al., 2015). We planned to stop collecting data after gathering usable data from 180 children (with approximately 60 children per age group). We ended up with a larger sample of children, however, for two reasons. First, we ultimately decided to adopt more inclusive criteria than we specified in the preregistration (because, in retrospect, we thought that the prespecified criteria were too stringent for the youngest children; this allowed an additional 44 children to be retained for analyses). Second, we originally thought that data from 57 participants were lost because of technical difficulties, but these data were ultimately recoverable. Analyses that precisely followed our preregistered criteria and stopping rule (e.g., analyses of the first 180 children who were recruited who passed the prespecified inclusion criteria) revealed similar patterns as reported here (see https://osf.io/64yx5/).

Forty-three additional children and four adults participated in the study but were excluded for not completing the entire study (n = 7 children), incorrectly responding to attention checks (n = 8 children, 4 adults), interference from parents (n = 2 children), or technical difficulties with the testing software (e.g., software failed to record data in a manner that was recoverable; n = 26 children).

Procedure. Participants completed two tasks in a fixed order: (a) a novel-category inference task and (b) a context-dependent pragmatic-ability task previously used with children even younger than those included here (Stiller et al., 2015). We presented these tasks on touch-screen tablets using http://testable.org. The context-dependent pragmatic-ability task was included to make sure that this testing environment (using animations on a screen) could successfully elicit pragmatic reasoning in young children.[1] An example of the full procedure and materials can be found at https://nyu-cdsc.github.io/pidi-demo/.

Novel-category inference task.

Learning phase. A recorded narrator introduced participants to a special town that had only two kinds of people: zarpies and gorps (see Fig. 1 for a summary of the protocol). Group membership was perceptually marked by the clothing color (e.g., zarpies wore green, and gorps wore yellow; group color was randomized). Besides wearing the same clothing color, members of both groups were otherwise diverse with respect to race/ethnicity, gender, and physical features. Next, to introduce the groups and make sure that children understood them as meaningful and distinct, the narrator told participants four generic statements about each group (eight statements in total) in alternating order (e.g., "Zarpies like to climb tall fences," "Gorps like to draw stars on their knees"). The learning phase included two knowledge checks—one in which participants had to verbally provide the correct label for each group and one in which they had to point to members of each group—to ensure that participants learned the two groups. Participants were given feedback on their responses regardless of whether they answered the question correctly and, overall, had a mean accuracy level of 91% (only five children answered both questions incorrectly).

Test phase. Following the learning phase, participants were presented with four test trials. For each trial, the narrator introduced participants to a new group of zarpies and gorps. The narrator then directed the participant's attention to an individual (i.e., "Look at this zarpie!") and described a property (e.g., "good at baking pizzas") true of that individual. The language used to describe the property varied by condition: Participants in the *generic* condition heard about this property via a generic statement (e.g., "Zarpies are good at baking pizza"), whereas participants in the *specific* condition learned about this property via a specific statement (e.g., "This zarpie is good at baking pizza"). The language used to describe the target property across the four test trials is the only way in which the two conditions (generic vs. specific) differed from one another.

After the target property was introduced, the individual moved to the top right corner of the screen to serve as a memory cue, and the language used to introduce the target property was repeated once more (e.g., "I'll move it here to remind you that zarpies are good at baking pizzas" for the generic condition). Finally, participants responded either yes or no regarding whether two additional individuals—a zarpie and a gorp drawn from a group of zarpies and gorps—also

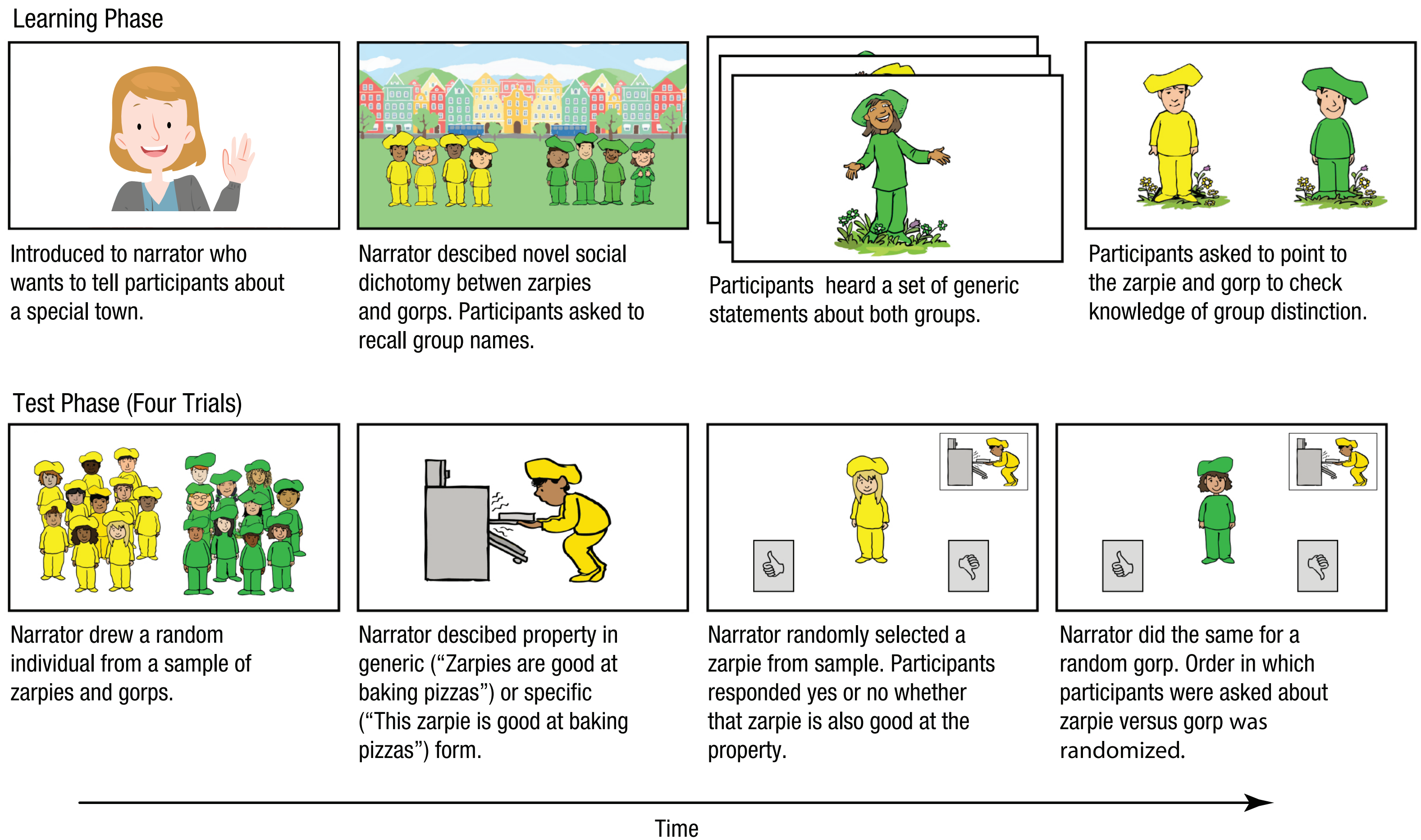

Fig. 1. Summary of the protocol for the learning and test phases of the novel-category inference task used in Study 1 (with sample stimuli). Zarpies are shown here in yellow and gorps in green. Study 2 used the exact same protocol, except that participants did not hear the set of generic statements in the third panel in the top row. The leftmost panel in the top row was adapted from resources found on Freepik.com.

possessed the property (e.g., whether the new zarpie/gorp was also good at baking pizzas). Participants were asked about the individuals one at a time, and the order of the two questions (i.e., whether participants were asked about a zarpie or gorp first) was randomized across trials. Responses to these two questions across the four trials were used in our analyses.

Results

Our data were compiled, processed, and analyzed in the R programming environment (Version 3.6.1; R Core Team, 2019) using the following packages: *tidyverse* (Wickham et al., 2019), *geepack* (Halekoh, Højsgaard, & Yan, 2006), *lubridate* (Grolemund & Wickham, 2011), *here* (Müller, 2017), *interactions* (Long, 2019), *lme4* (Bates, Mächler, Bolker, & Walker, 2015), *emmeans* (Lenth, 2020), *cowplot* (Wilke, 2020), and *ggpubr* (Kassambara, 2020). Our R code is available on OSF (https://osf.io/3um4k).

We analyzed participants' choices on the novel-category inference task using generalized estimating equations (GEEs) because they can assess both within-subjects and between-subjects effects in binary data. GEEs yield Wald χ^2 values as indicators of main effects and interactions. The models tested here used a binomial outcome distribution with a logit link function and an exchangeable correlation matrix (Ballinger, 2004). For analyses including adults, we treated age as a categorical variable (4-year-olds, 5-year-olds, 6-year-olds, and adults); for comparable analyses with children only, we used children's exact age (e.g., 4.53 years), treated as a continuous variable. We report the results of the Wald χ^2 tests as well as estimated marginal means and slopes with 95% confidence intervals (CIs).

Property extension. First, we examined the likelihood of extending the target property (i.e., responding "yes") to other individuals as a function of the individual's group membership (from previously mentioned or unmentioned group), participant's age, and language condition (generic or specific; preregistered at https://osf.io/c3fpv/). Participants extended the target property more to members of the previously mentioned group (M = 69%, 95% CI = [66%, 73%]) than the unmentioned group (M = 28%, 95% CI = [25%, 31%]), group: $\chi^2(1, N = 343) = 158$, $p < .001$, but the tendency to do so was stronger after hearing generic claims (mentioned—generic: M = 80%, 95% CI = [75%, 84%] vs. specific: M = 60%, 95% CI = [55%, 65%]; unmentioned—generic: M = 18%, 95% CI = [14%, 22%] vs. specific: M = 37%, 95% CI = [32%, 42%]), Group × Condition: $\chi^2(1, N = 343) = 4.1$, $p < .001$, and with age, Age × Group: $\chi^2(3, N = 343) = 50.6$, $p < .001$; Age × Group × Condition: $\chi^2(3, N = 343) = 1.9$, $p = .008$ (see Fig. 2a). All other effects had ps greater than .10.

To tease apart the interaction among age, condition, and group, we used separate GEEs to analyze the simple effects of condition and stimulus for each age group. At each age, participants responded that the individual from the unmentioned group did not have the property more often if they heard the property described with generic rather than specific language, Condition × Group—4-year-olds: $\chi^2(1, N = 112) = 9.86$, $p = .002$; 5-year-olds: $\chi^2(1, N = 83) = 9.05$, $p = .003$; 6-year-olds: $\chi^2(1, N = 92) = 25.5$, $p < .001$; adults: $\chi^2(1, N = 56) = 13.7$, $p < .001$.

These patterns of results held when we examined only child participant data (age as continuous), including the three-way interaction among age, condition, and group membership, $\chi^2(1, N = 285) = 9.88$, $p = .002$. With age, children were more likely to extend the target property to other individuals from the mentioned group (β = 0.62, 95% CI = [0.28, 0.95]) and less likely to extend the target property to other members of the unmentioned group (β = −0.68, 95% CI = [−1.10, −0.26]) after hearing generic language. Their rate of extending the target property after hearing specific language—to either the mentioned group (β = 0.12, 95% CI = [−0.16, 0.40]) or the unmentioned group (β = 0.0007, 95% CI = [−0.28, 0.28])—did not shift with age. Critically, by 4.7 years of age (determined using the Johnson-Neyman technique; Johnson & Fay, 1950), children extended the target property to members of the unmentioned group less often after hearing generic language than specific language (see Table 1).

Property inference. In the previous analyses, we examined the likelihood that participants extended the target property to mentioned and unmentioned groups, averaged across all trials. After submitting our preregistration, we realized that an even stronger test of the hypothesis that participants infer that properties expressed via generic claims are not true of unmentioned groups would be to examine participants' responses on a trial-by-trial basis. That is, for a given target property, do participants infer both (a) that the property is true (i.e., respond "yes") of members of the mentioned group but (b) that the property is not true (i.e., respond "no") of members of the unmentioned group? Participants who responded in this pattern were marked as making the inference and assigned a 1 for that trial; participants who responded in any other pattern did not make the expected inference and were assigned a 0.

When examining both the child and adult data, we found that participants were more likely to make the expected inference after hearing a generic statement

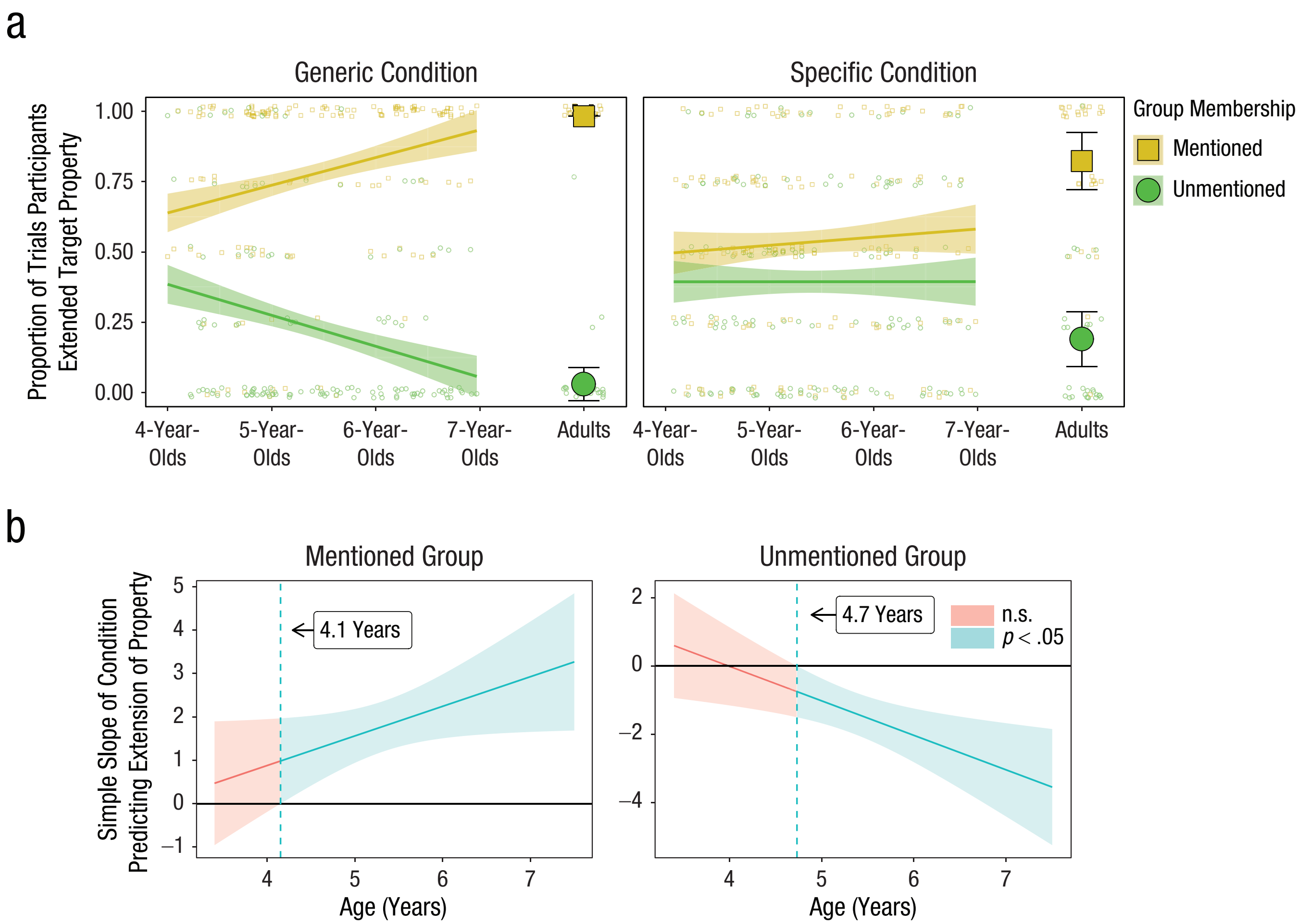

Fig. 2. Property extension in Study 1. The proportion of trials in which participants extended the target property to individuals of the mentioned and unmentioned groups (a) is shown separately for each age and condition. Error bars and bands reflect 95% confidence intervals. Small shapes reflect individual averages, and large shapes and lines reflect group averages. In the Johnson-Neyman plots (b), lines reflect differences in simple slopes of condition predicting participants' extension of the target property as a function of age and group membership. The age at which property extension became significantly different is marked by the dashed vertical line.

(generic: M = 69%, 95% CI = [66%, 73%]; specific: M = 41%, 95% CI = [38%, 45%]), condition: $\chi^2(1, N = 343) = 42.8$, $p < .001$, and the tendency to do so increased with age—age: $\chi^2(3, N = 343) = 43.3$, $p < .001$; Age × Condition: $\chi^2(3, N = 343) = 8.89$, $p = .031$ (see Fig. 3a). Importantly, across all four age groups, participants who heard generic statements made the expected inference (i.e., responding "yes" to a member of the mentioned group but "no" to a member of the unmentioned group) more often than participants who heard specific statements (4-year-olds: $p = .002$; 5-year-olds: $p = .013$; 6-year-olds: $p < .001$; adults: $p = .002$).

When we examined the child data alone (with age treated continuously), the two-way interaction between age and condition remained significant, $\chi^2(1, N = 285) = 8.33$, $p = .004$. Using the Johnson-Neyman procedure (Johnson & Fay, 1950), we found that children older than 4.5 years made the expected inference significantly more frequently after hearing generic language than specific language (see Table 1), and the tendency to make inferences from generic claims increased with age (generic: β = 0.82, 95% CI = [0.45, 1.19]; specific: β = 0.14, 95% CI = [−0.14, 0.42]).

Together, these findings suggest that children as young as 4.5 years infer meaning about unmentioned groups from generic claims.

Study 2

Study 2 was designed to replicate Study 1 with one key change. In Study 1's learning phase, children (in both language conditions) heard generic claims about both categories. Although this was intended to help children learn the categories, it raised the possibility

Table 1. Average Percentage of Trials in Which Participants Extended the Target Property and Made the Expected Inference in Study 1

Condition and group	Four-year-olds	Five-year-olds	Six-year-olds	Adults
		Property extension		
Generic				
Mentioned	73% [65%, 80%]	73% [65%, 81%]	89% [82%, 96%]	100% [98%, 100%]
Unmentioned	32% [25%, 39%]	19% [11%, 27%]	14% [7%, 21%]	5% [0%, 9%]
Specific				
Mentioned	53% [45%, 61%]	51% [42%, 60%]	59% [50%, 67%]	81% [72%, 90%]
Unmentioned	41% [33%, 49%]	34% [26%, 43%]	42% [33%, 51%]	18% [8%, 27%]
		Property inference		
Generic	52% [46%, 59%]	61% [54%, 69%]	82% [76%, 87%]	95% [91%, 99%]
Specific	32% [25%, 38%]	40% [33%, 47%]	39% [32%, 46%]	66% [57%, 74%]

Note: Values in brackets are 95% confidence intervals.

that children would make inferences about unmentioned categories only after first hearing a series of distinct generic claims about both groups. By removing this from the introduction, Study 2 tested whether children's inferences from generic claims rely on a cumulative effect after hearing many generic claims or whether these inferences persist even after hearing only a single generic claim with minimal introduction to each group.

Method

We followed the methodological and analytic decisions preregistered for Study 1. Although we found age-related

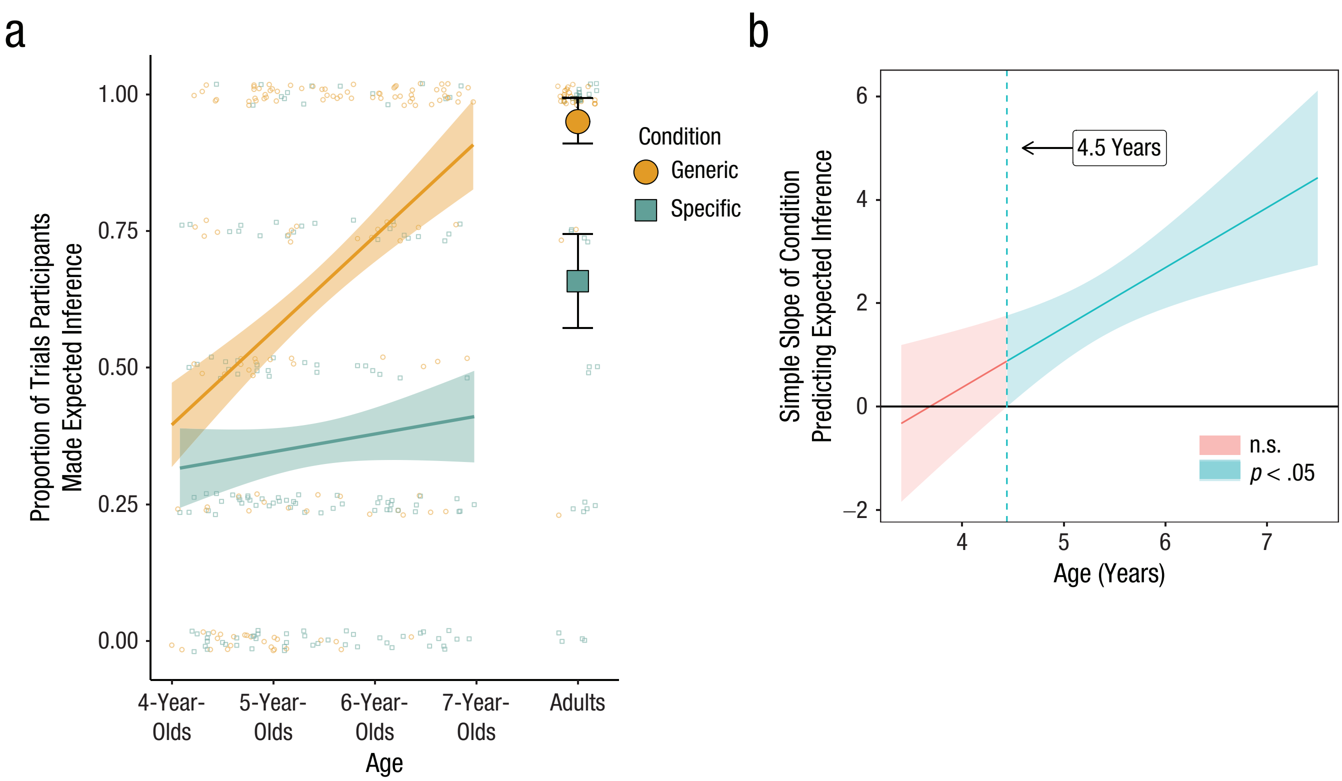

Fig. 3. Property inference in Study 1. The proportion of trials in which participants made the expected inference (i.e., responded "yes" that an individual of the mentioned group had the target property but "no" that an individual of the unmentioned group did not; a) is shown as a function of age and condition. Error bars and bands reflect 95% confidence intervals. Small shapes reflect individual averages, and large shapes and lines reflect group averages. In the Johnson-Neyman plot (b), the line reflects a difference in the simple slope of condition predicting participants' inferences as a function of age. The age at which making inferences across conditions became significantly different is marked by the dashed vertical line.

Table 2. Average Percentage of Trials in Which Participants Extended the Target Property and Made the Expected Inference in Study 3

Condition and group	Four-year-olds	Five-year-olds	Six-year-olds	Adults
		Property extension		
Knowledge				
Mentioned	80% [69%, 90%]	78% [67%, 88%]	85% [78%, 93%]	99% [94%, 100%]
Unmentioned	46% [37%, 56%]	33% [11%, 27%]	15% [7%, 23%]	15% [9%, 22%]
No knowledge				
Mentioned	54% [41%, 66%]	52% [40%, 63%]	75% [64%, 86%]	84% [75%, 92%]
Unmentioned	50% [37%, 63%]	46% [35%, 58%]	55% [44%, 66%]	41% [31%, 52%]
		Property inference		
Knowledge	51% [42%, 59%]	60% [51%, 69%]	79% [72%, 86%]	83% [77%, 89%]
No knowledge	33% [24%, 42%]	34% [25%, 43%]	38% [30%, 47%]	52% [43%, 61%]

Note: Values in brackets are 95% confidence intervals.

changes in Study 1 (which we explored in more detail in Study 3), for this follow-up, we did not intend to examine age-related changes and considered children as a single group.

Participants. Eighty-four children (62% girls), ages 4 to 6 years (31 four-year-olds, M = 4.53 years; 25 five-year-olds, M = 5.57 years; 28 six-year-olds, M = 6.52 years), participated in this study. Of those parents who provided racial and ethnic demographic information for their children (n = 68), 59% identified their child as White, 15% as Asian, 9% as Black/African American, and 18% as mixed or biracial; 21% of these participants were also Hispanic/Latino. Children were recruited and assigned to a condition (generic vs. specific) in the same manner as in Study 1. Additionally, as in Study 1, we determined our sample size on the basis of comparable studies exploring the development of pragmatic reasoning (e.g., Horowitz & Frank, 2016; Stiller et al., 2015). Fifteen additional children participated but were excluded for technical difficulties with the testing software (n = 10), failure to complete the entire study (n = 1), and failure to correctly answer attention checks (n = 4).

Procedure. Procedures were identical to those in Study 1, except that in the novel-category inference task, we removed the portion of the learning phase in which children heard generic language about both groups. That is, in Figure 1, we removed the portion depicted in the third panel in the top row.

Results

As in Study 1, we analyzed participants' choices on the novel-category inference task using GEEs. We followed the same strategy of analyses as in Study 1, first examining how children extended the target property and then examining what inferences children made. Given the smaller sample size, we did not examine age-related changes in children's responses. We report the results of the Wald χ^2 tests and estimated marginal means with 95% CIs.

Property extension. For property extension, we found the same pattern of results as in Study 1: Children were more likely to extend the target property to members of the previously mentioned group (M = .68, 95% CI = [.60, .75]) than the unmentioned group (M = .33, 95% CI = [.26, .40]), group: $\chi^2(1, N = 84) = 30.8$, $p < .001$, but whether children did so depended on the language used to introduce the target property, Group × Condition: $\chi^2(1, N = 84) = 7.41$, $p = .006$ (see Fig. 4a). Children who heard generic claims extended the target property to members of the mentioned group (M = 75%, 95% CI = [67%, 84%]) more frequently than to members of the unmentioned group (M = 25%, 95% CI = [16%, 34%], p = .007). However, children who heard specific language did not differentiate their property extensions to members of the mentioned group (M = 60%, 95% CI = [50%, 69%]) and unmentioned group (M = 41%, 95% CI = [31%, 50%], p = .12).

Property inference. Next, we examined whether, at the level of individual trials, children inferred that the target property was true of the mentioned group but not of the unmentioned group. We found that children more frequently made this inference after hearing the property conveyed via a generic statement (M = 66%, 95% CI = [55%, 75%]) than via a specific statement (M = 40%, 95% CI = [30%, 51%]), $\chi^2(1, N = 84) = 10.6$, p = .001 (see Fig. 4b), thus replicating and extending the findings from Study 1 by demonstrating that children's inferences from generic claims do not depend on a cumulative effect from many generic claims but rather persist even with minimal introduction to both groups.

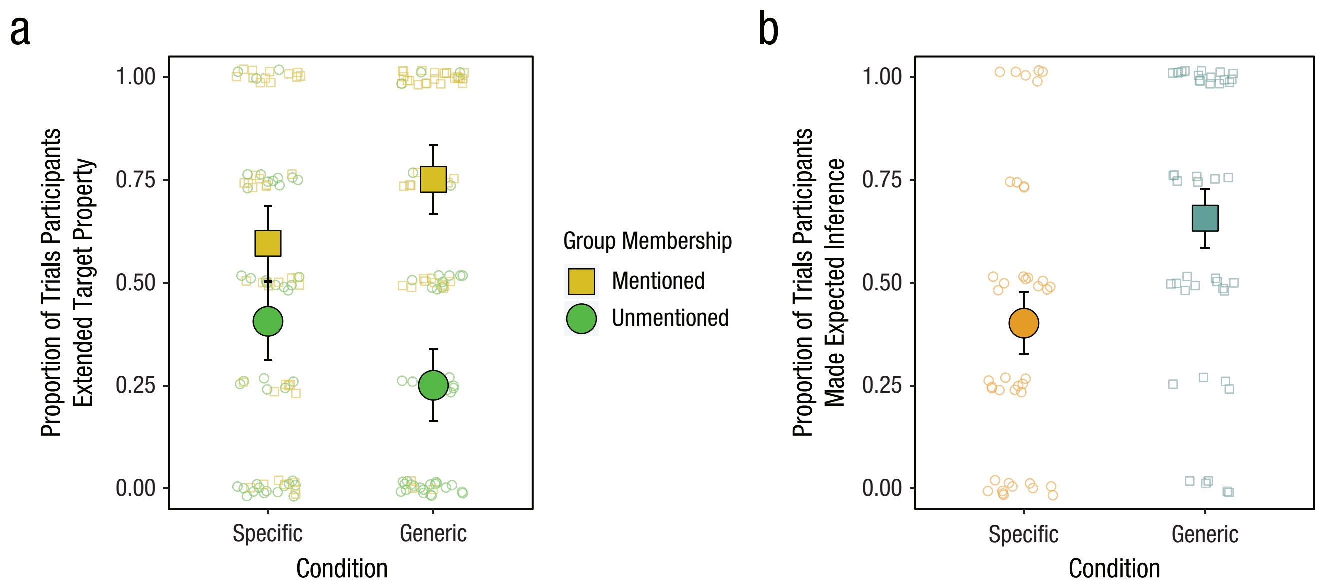

Fig. 4. Property extension in Study 2. The proportion of trials in which children extended the target property to individuals of the mentioned and unmentioned groups (a) is shown for each condition. The proportion of trials in which children made the expected inference (b) is shown for each condition. Error bars reflect 95% confidence intervals. Small shapes reflect individual averages, and large shapes reflect group averages.

Study 3

Studies 1 and 2 demonstrated that children make inferences about unmentioned groups from generic claims. Study 3 provided a direct test of the mechanism underlying these inferences: Do children generate these inferences by reasoning pragmatically about the mental states of the speaker, or do they do so using some kind of heuristic that does not employ higher level reasoning about mental states (e.g., applying a rule that contrasting groups are opposite)? If pragmatic ability underlies these inferences, then children should infer meaning about unmentioned groups only when the generic claim is communicated by someone knowledgeable.

Method

Preregistered hypotheses. We predicted that participants who heard, "Zarpies are good at baking pizzas," for example, from a knowledgeable speaker (but not from an unknowledgeable one) would infer that gorps are not good at baking pizzas. Given Study 1's findings, we hypothesized that there would be an age-related increase in children's tendency to infer meaning about unmentioned groups from generic claims conveyed by knowledgeable speakers. The procedures and analytic code to test these hypotheses were preregistered and can be found at https://osf.io/ykwg4 (a preregistered addendum can be found at https://osf.io/wfxrv).

Participants. One hundred eighty-one children (50% girls), ages 4 to 7.25 years (59 four-year-olds, M = 4.60 years; 56 five-year-olds, M = 5.47 years; 66 six- to 7.25-year-olds, M = 6.55 years), and 65 adults (M = 34.8 years, 22% women) participated in this study. Of those parents who provided racial and ethnic demographic information for their children (n = 161), 56% of them identified their child as White, 20% as Asian, 7% as Black/African American, 3% as Middle Eastern or North African, and 13% as mixed or biracial; 22% of these participants were also Hispanic/Latino. We did not gather demographic data for adult participants; however, participants were likely predominantly White, as was the case in Study 1 and in other studies (e.g., Huff & Tingley, 2015) that have examined the racial and ethnic makeup of participants recruited via Amazon's Mechanical Turk.

Children and adults were recruited in the same manner as in Studies 1 and 2. As was the case in Study 1, we preregistered a sample size of 60 participants per age bin (4-year-olds, 5-year-olds, 6-year-olds, and adults) with half in each condition (planned total N = 240 [180 children, 60 adults]) on the basis of the sample sizes used in comparable studies exploring pragmatic ability in early childhood (e.g., Horowitz & Frank, 2016; Stiller et al., 2015). We planned to stop data collection after gathering usable data from 180 children (with ~60 children per age group). Approximately half of participants were assigned to one of two conditions (knowledgeable speaker vs. unknowledgeable speaker). One additional adult participated in the study but was excluded for incorrectly responding to attention checks. Eight children were excluded for failure to complete the entire task ($n = 2$), incorrectly responding to attention checks ($n = 5$), or technical difficulties ($n = 1$).

Procedure. Children participants completed three tasks in the following order: a warm-up task, a novel-category inference task, and a context-dependent pragmatic-ability task. Adult participants completed only the warm-up task and novel-category inference task. We presented these tasks using an open-source application for running behavioral experiments (https://github.com/nyu-cdsc/toku) developed by the first author. An example of the procedure can be found at https://nyu-cdsc.github.io/pidi-demo/, and materials are available at https://osf.io/m4jwf/.

Novel-category inference task. This task was similar to those used in Studies 1 and 2. We describe the differences in Study 3 below.

Learning phase. As in Studies 1 and 2, a narrator introduced participants to a special town that has only two kinds of people (zarpies and gorps). However, in Study 3, half of the participants learned about the novel social dichotomy from a knowledgeable speaker (i.e., the speaker lived in the same neighborhood), and the other half learned about the dichotomy from an unknowledgeable speaker (i.e., the speaker was visiting the town for the first time; see https://osf.io/8aqfc/ for scripts). Participants were then asked a set of knowledge checks to ensure that they were (a) familiar with the distinction between zarpies and gorps and (b) aware of the speaker's knowledge or lack thereof of the two groups. Participants received feedback on their responses to these questions to ensure that they learned the category boundaries and critical manipulation. Children were not excluded for incorrectly answering these questions and performed well overall (M = 80%; 83% of children answered at least three of the four questions correctly, and no child answered all four questions incorrectly).

Test phase. Participants saw four test trials similar to those used in Studies 1 and 2, with two notable differences. First, regardless of condition (knowledgeable speaker vs. unknowledgeable speaker), the speaker saw a category member doing the target property and then introduced the property using a generic statement (e.g., speakers in both conditions saw an individual zarpie baking pizzas and then stated, "Zarpies are good at baking pizzas"). Thus, speakers in both conditions had access to the same evidence (of an individual zarpie baking a pizza) at the time they made their claim; what differed across condition was the background knowledge about zarpies and gorps that the speaker brought to the task. Halfway through, participants received a reminder about the speaker's background knowledge (i.e., whether this was their first time visiting the town). Second, in previous versions of the task, during the test phase, participants were given a visual cue about the target property, and the speaker's statement was repeated. In Study 3, participants did not receive this visual aid nor was the generic statement repeated. Removing these cues helped rule out possible alternative explanations that the findings from the previous studies were the result of low-level perceptual mechanisms.

Results

Property extension. Following the same analysis strategy described in Study 1, we first examined the likelihood of extending the target property (i.e., responding "yes") to other individuals as a function of the individual's group membership (from previously mentioned or unmentioned group), participant's age, and condition (whether the speaker was knowledgeable or unknowledgeable). Children and adults were more likely to extend the target property to members of the previously mentioned group than the unmentioned group, group: $\chi^2(1, N = 246) = 123, p < .001$. Critically, their tendency to do so depended on the speaker's knowledge of two groups, Group × Condition: $\chi^2(1, N = 246) = 39.5, p < .001$, and the participant's age, Age × Group: $\chi^2(3, N = 246) = 50.1, p < .001$; Age × Condition: $\chi^2(3, N = 246) = 18.9, p < .001$; Age × Group × Condition: $\chi^2(3, N = 246) = 7.56, p = .056$ (see Fig. 5a). Across all four age groups, children and adults were more likely to infer that the target property was not true of individuals from the unmentioned group after hearing generic statements from a knowledgeable speaker compared with when they heard the exact same generic statement made by an unknowledgeable speaker, Group × Condition—4-year-olds: $\chi^2(1, N = 59) = 8.23, p = .004$; 5-year-olds: $\chi^2(1, N = 56) = 8.86, p = .003$; 6-year-olds: $\chi^2(1, N = 66) = 12.2, p < .001$; adults: $\chi^2(1, N = 65) = 20.6, p < .001$.

We found the same pattern of results when looking at children's responses only (these analyses used children's exact age as a continuous variable). Children inferred that the target property was true of individuals from the previously mentioned group but not true of individuals from the unmentioned group, but they did so only when the speaker was knowledgeable and the tendency to do increased with age, Age × Condition × Group: $\chi^2(1, N = 181) = 5.14, p = .023$. With age, children were less likely to extend the target property to members of the unmentioned group after hearing generic language from a knowledgeable speaker

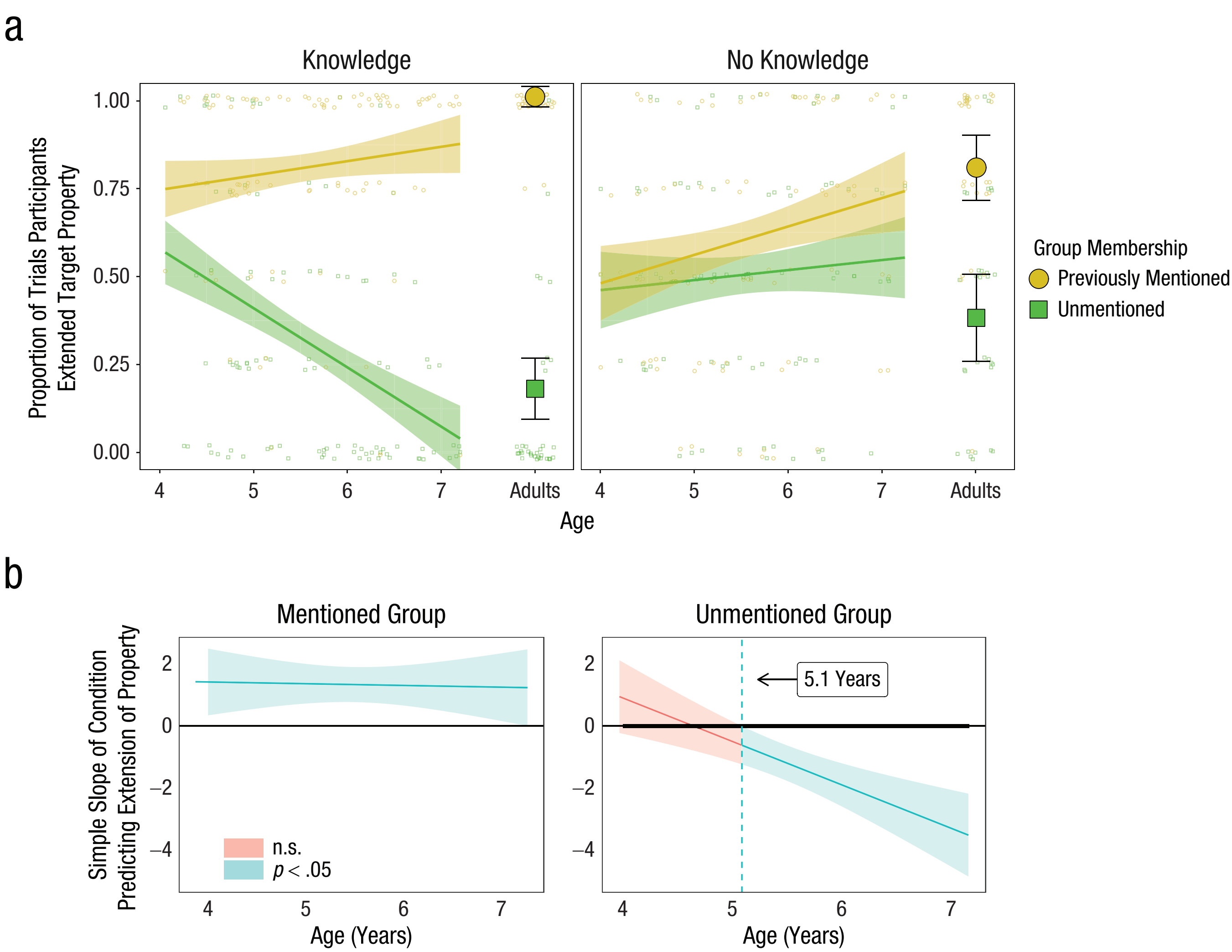

Fig. 5. Property extension in Study 3. The proportion of trials in which participants extended the target property to individuals of the mentioned and unmentioned groups (a) is shown separately for each age and condition. Error bars and bands reflect 95% confidence intervals. Small shapes reflect individual averages, and large shapes and lines reflect group averages. In the Johnson-Neyman plots (b), lines reflect differences in simple slopes of condition predicting participants' extension of the target property as a function of age and group membership. The age at which property extension to the unmentioned group became significantly different is marked by the dashed vertical line. The rate of property extension to the mentioned group was significantly different across all observed ages.

(β = −0.88, 95% CI = [−1.30, −0.45]) but not after hearing generic language from an unknowledgeable speaker (β = 0.12, 95% CI = [−0.16, −0.39]). Critically, by 5.1 years of age (determined using the Johnson-Neyman technique; Johnson & Fay, 1950; see Table 2), children extended the target property to members of the unmentioned group less often after hearing a generic claim from a knowledgeable speaker than from an unknowledgeable speaker.

Property inference. Next, looking at the inferences that participants made for a given target property (i.e., inferred that the mentioned group had the property but that the unmentioned group did not), we found that children and adults were more likely to make the expected inference when the generic claim was communicated by a knowledgeable speaker (M = .70, 95% CI = [.63, .76]) than by an unknowledgeable speaker (M = .39, 95% CI = [.33, .45]), condition: $\chi^2(1, N = 246) = 41.1$, $p < .001$ (see Fig. 6a). The tendency to make this kind of inference increased with age, age: $\chi^2(3, N = 246) = 24.1$, $p < .001$. Although we did not find an interaction between age (when treated categorically) and condition with the adult data included in the model, Age × Condition: $\chi^2(3, N = 246) = 4.81$, $p = .19$, we did see this interaction between age and condition when examining the child data alone (here, age was treated continuously), Age × Condition: $\chi^2(1, N = 181) = 8.16$, $p = .004$. By 4.7 years of age (determined using the Johnson-Neyman procedure; Johnson & Fay, 1950; see Table 2), children were more likely to infer that properties

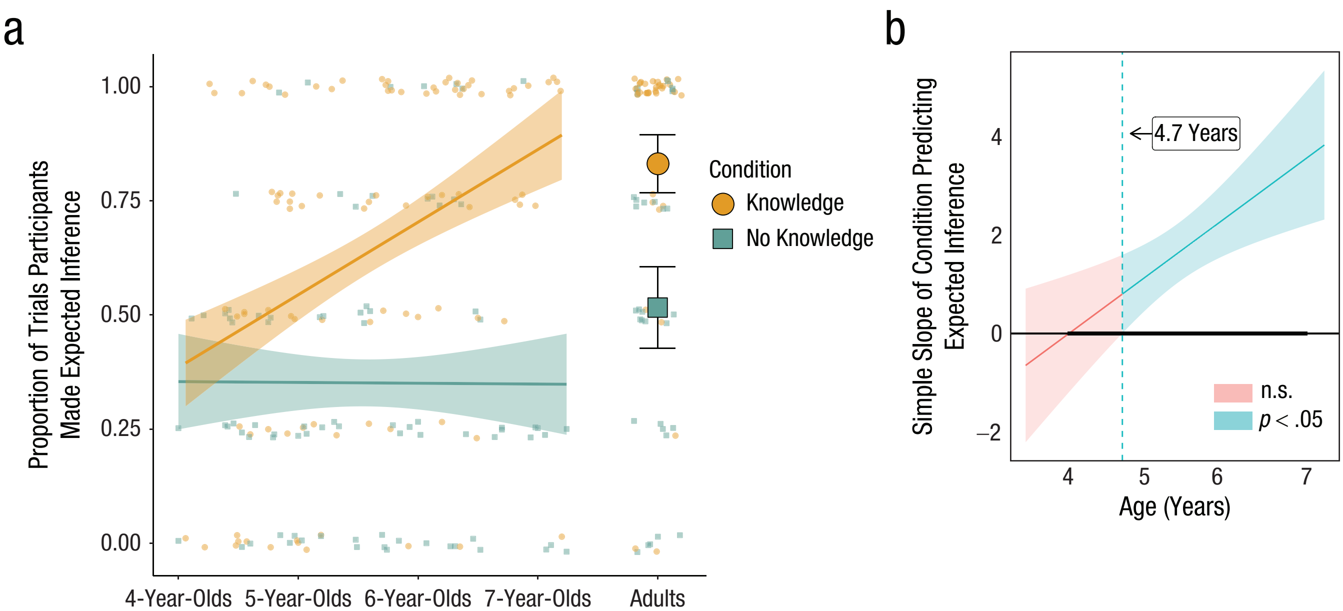

Fig. 6. Property inference in Study 3. The proportion of trials in which participants made the expected inference (i.e., responded "yes" that an individual of the mentioned group had the target property but "no" that an individual of the unmentioned group did not; a) is shown as a function of age and condition. Error bars and bands reflect 95% confidence intervals. Small shapes reflect individual averages, and large shapes and lines reflect group averages. In the Johnson-Neyman plot (b), the line reflects a difference in slopes between knowledgeable and unknowledgeable speaker conditions across age. The rate at which making inferences across conditions became significantly different is marked by the dashed vertical line.

conveyed via generic statements are not true of unmentioned groups when the generic statement was made by a knowledgeable speaker than when the same generic statement was said by an unknowledgeable speaker. The tendency to make inferences about unmentioned groups from generic statements only when conveyed by knowledgeable speakers increased with age ($\beta = 0.74$, 95% CI = [0.32, 1.17]).

Together, these findings suggest that children engage in pragmatic reasoning (as opposed to employing a general heuristic or rule) when making inferences about unmentioned groups, given that they do so only when the generic statement is communicated by a knowledgeable speaker.

General Discussion

In these three studies, children made inferences about unmentioned categories from generic statements. That is, after hearing "Zarpies are good at baking pizzas," they assumed not only that a new zarpie would be good at this but also that a new gorp would not. Children made these inferences only when generic statements were said by knowledgeable speakers, suggesting that children reason about a speaker's intended meaning to draw these inferences. Children's tendency to make these inferences emerged early (around 4.5 years) and became more robust with age.

Beyond providing insight on children's inferences from generic statements, these studies also provide evidence for a potential mechanism by which social stereotypes may be inadvertently communicated to young children. Parents may believe that generic statements—such as "Boys are good at sports" or "Girls play with dolls"—provide an innocuous way to describe social groups. In fact, parents use generic language frequently when talking with their children (e.g., Gelman et al., 2008). But these studies suggest that young children take away not only the positive message from these statements (e.g., that boys are good at sports) but also an additional negative message about other groups not explicitly mentioned (e.g., that girls are not good at sports). This can reinforce existing stereotypes or perhaps even create new ones. This is likely especially true when social categories have clear and salient contrasts, as was the case in the present studies as well as with gender (when presented as binary: boys vs. girls).

This idea that generic statements may communicate stereotypes about unmentioned groups builds on a growing body of research demonstrating the power of generic statements in transmitting social stereotypes about explicitly mentioned groups to young children—by signaling how to carve up the social world into meaningful categories (Rhodes, Leslie, Bianchi, & Chalik, 2018), promoting essentialist beliefs about social groups

(Rhodes et al., 2012), and consequently increasing social stereotyping and negative intergroup relations (e.g., Hammond & Cimpian, 2017; Leslie, 2017; Rhodes, Leslie, Saunders, Dunham, & Cimpian, 2018). Although we primarily focused on children, generic statements also bias the beliefs of adults (as confirmed here, among other studies, e.g., Cimpian et al., 2010; Rhodes et al., 2012) and have broad consequences for reasoning across domains, including how they interpret and value scientific research (DeJesus, Callanan, Solis, & Gelman, 2019).

The present studies suggest a number of interesting open questions that future research should explore. First, it would be informative to examine the developmental mechanisms underlying the age-related increases in the robustness of children's inferences. For instance, these inferences could become more robust because of increased efficiency with pragmatic reasoning generally or because of more general improvements in children's ability to make sense of our experimental task. Consistent with this latter possibility, the results of our supplemental analyses, which included only children who met the more stringent inclusion criteria of passing all attention checks, showed that children reliably made these inferences about 6 months earlier than was found in analyses with the broader sample. Second, it is unknown whether the unmentioned group needs to be physically present to warrant the inference (as was the case in the present studies) or whether children would also extend inferences about unmentioned groups even when they are not perceptually available. The latter may be particularly difficult for young children, who already struggle to determine the relevant scope for pragmatic inferences (e.g., Barner et al., 2011). Third, these studies suggest that children make inferences about unmentioned groups when the mentioned and unmentioned groups contrast along a social dichotomy (akin to a binary and discrete perspective of gender). However, many social categories have multiple relevant contrasts, as is the case for race, or are presented as having less distinct or discrete boundaries. Future research should examine the types of inferences that children draw about unmentioned groups in these cases, as well as how these inferences play out for real-world categories (e.g., gender-based inferences). In sum, the present studies highlight how implicit features of communication can shape the development of children's social understanding, paving the way for future work to examine the full extent of these effects.

Transparency

Action Editor: Bill von Hippel
Editor: D. Stephen Lindsay
Author Contributions
K. Moty developed the study concept. Both authors contributed to the study design. K. Moty conducted testing and data collection. K. Moty analyzed and interpreted the data under the supervision of M. Rhodes. K. Moty drafted the manuscript, and M. Rhodes provided critical revisions. Both authors approved the final manuscript for submission.
Declaration of Conflicting Interests
The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.
Funding
This research was supported by National Institute of Child Health and Human Development Grant R01HD087672 and National Science Foundation Grant 1729540 (to M. Rhodes) and by a National Science Foundation Graduate Research Fellowship (to K. Moty). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or National Science Foundation.
Open Practices
All data and analysis code have been made publicly available via OSF and can be accessed at https://osf.io/3um4k. All materials have been made publicly available at https://osf.io/m4jwf/. Studies 1 and 2 were preregistered at https://osf.io/c3fpv, and Study 3 was preregistered at https://osf.io/ykwg4 (an addendum to Study 3 was preregistered at https://osf.io/wfxrv). Deviations from the preregistrations are discussed in the text. This article has received the badges for Open Data, Open Materials, and Preregistration. More information about the Open Practices badges can be found at http://www.psychologicalscience.org/publications/badges.
TC

ORCID iD

Kelsey Moty https://orcid.org/0000-0002-8580-0896

Acknowledgments

We thank Angela Sorenson, Alison De Leon, Jerolynn Liu, Kevin Bui, Samantha Aboody, Samuel Krystal, and Shelly Reizen for assistance with data collection. We also thank the staff and families at the Children's Museum of Manhattan, New York City District 2 preschools, and Public School 132 for their support and participation.

Note

1. Performance on the pragmatic-ability task across all three studies was excellent (near ceiling) and will not be discussed further in the article (see https://osf.io/64yx5/).

References

Ballinger, G. A. (2004). Using generalized estimating equations for longitudinal data analysis. *Organizational Research Methods*, *7*, 127–150. doi:10.1177/1094428104263672

Barner, D., Brooks, N., & Bale, A. (2011). Accessing the unsaid: The role of scalar alternatives in children's pragmatic inference. *Cognition*, *118*, 84–93. doi:10.1016/j.cognition.2010.10.010

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). doi:10.18637/jss.v067.i01

Brandone, A., Gelman, S. A., & Hedglen, J. (2015). Children's developing intuitions about the truth conditions and implications of novel generics versus quantified statements. *Cognitive Science*, *39*, 711–738. doi:10.1111/cogs.12176

Carlson, G. N. (1977). A unified analysis of the English bare plural. *Linguistics and Philosophy*, *1*, 413–457. doi:10.1007/BF00353456

Carlson, G. N., & Pelletier, F. J. (1995). *The generic book*. Chicago, IL: University of Chicago 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. doi:10.1111/j.1551-6709.2010.01126.x

Cimpian, A., & Erickson, L. C. (2012). Remembering kinds: New evidence that categories are privileged in children's thinking. *Cognitive Psychology*, *64*, 161–185. doi:10.1016/j.cogpsych.2011.11.002

Cimpian, A., & Markman, E. M. (2009). Information learned from generic language becomes central to children's biological concepts: Evidence from their open-ended explanations. *Cognition*, *113*, 14–25. doi:10.1016/j.cognition.2009.07.004

Cimpian, A., & Markman, E. M. (2011). The generic/nongeneric distinction influences how children interpret new information about social others. *Child Development*, *82*, 471–492. doi:10.1111/j.1467-8624.2010.01525.x

Cimpian, A., & Salomon, E. (2014). The inherence heuristic: An intuitive means of making sense of the world, and a potential precursor to psychological essentialism. *Behavioral and Brain Sciences*, *37*, 461–480. doi:10.1017/s0140525x13002197

Clark, H. (1996). *Using language*. Cambridge, England: Cambridge University Press. doi:10.1017/CBO9780511620539

Cox, A. (2004, May 3). U.S. poised for epidemic West Nile year. *CNN*. Retrieved from http://edition.cnn.com/2004/HEALTH/05/03/wnv.outlook/

DeJesus, J. M., Callanan, M. A., Solis, G., & Gelman, S. A. (2019). Generic language in scientific communication. *Proceedings of the National Academy of Sciences, USA*, *116*, 18370–18377. doi:10.1073/pnas.1817706116

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. doi:10.1080/15475440701542625

Gelman, S. A., Taylor, M. G., & Nguyen, S. P. (2004). Mother-child conversations about gender: Understanding the acquisition of essentialist beliefs: IV. Talk about categories versus individuals (generics vs. non-generics). *Monographs of the Society for Research in Child Development*, *69*(1), 64–75. doi:10.1111/j.1540-5834.2004.06901005.x

Gelman, S. A., Ware, E. A., & Kleinberg, F. (2010). Effects of generic language on category content and structure. *Cognitive Psychology*, *61*, 273–301. doi:10.1016/j.cogpsych.2010.06.001

Graham, S. A., Gelman, S. A., & Clarke, J. (2016). Generics license 30-month-olds' inferences about the atypical properties of novel kinds. *Developmental Psychology*, *52*, 1353–1362. doi:10.1037/dev0000183

Graham, S. A., Nayer, S. L., & Gelman, S. A. (2011). Two-year-olds use the generic/nongeneric distinction to guide their inferences about novel kinds. *Child Development*, *82*, 493–507. doi:10.1111/j.1467-8624.2010.01572.x

Grice, H. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics* (Vol. 3, pp. 41–58). San Diego, CA: Academic Press.

Grolemund, G. & Wickham, H. (2011). Dates and times made easy with lubridate. *Journal of Statistical Software*, *40*(3). doi:10.18637/jss.v040.i03

Halekoh, U., Højsgaard, S., & Yan, J. (2006). The R package geepack for generalized estimating equations. *Journal of Statistical Software*, *15*(2). doi:10.18637/jss.v015.i02

Hammond, M. D., & Cimpian, A. (2017). Investigating the cognitive structure of stereotypes: Generic beliefs about groups predict social judgments better than statistical beliefs. *Journal of Experimental Psychology: General*, *146*, 607–614. doi:10.1037/xge0000297

Horn, L. (1984). Toward a new taxonomy for pragmatic inference: Q-based and r-based implicature. In D. Schiffrin (Ed.), *Meaning, form, and use in context: Linguistic applications* (pp. 11–42). Washington, DC: Georgetown University Press.

Horowitz, A. C., & Frank, M. C. (2016). Children's pragmatic inferences as a route for learning about the world. *Child Development*, *87*, 807–819. doi:10.1111/cdev.12527

Huang, Y. T., & Snedeker, J. (2009). Semantic meaning and pragmatic interpretation in 5-year-olds: Evidence from real-time spoken language comprehension. *Developmental Psychology*, *45*, 1723–1739. doi:10.1037/a0016704

Huff, C., & Tingley, D. (2015). "Who are these people?" Evaluating the demographic characteristics and political preferences of MTurk survey respondents. *Research & Politics*, *2*(3). doi:10.1177/2053168015604648

Jara-Ettinger, J., Floyd, S., Huey, H., Tenenbaum, J. B., & Schulz, L. E. (2020). Social pragmatics: Preschoolers rely on commonsense psychology to resolve referential underspecification. *Child Development*, *91*, 1135–1149. doi:10.1111/cdev.13290

Jaswal, V. K., & Neely, L. A. (2006). Adults don't always know best: Preschoolers use past reliability over age when learning new words. *Psychological Science*, *17*, 757–758. doi:10.1111/j.1467-9280.2006.01778.x

Johnson, P. O., & Fay, L. C. (1950). The Johnson-Neyman technique, its theory and application. *Psychometrika*, *15*, 349–367. doi:10.1007/bf02288864

Kassambara, A. (2020). ggpubr: 'ggplot2' based publication ready plots (R Package Version 0.4.0) [Computer software]. Retrieved from https://CRAN.R-project.org/package=ggpubr

Koenig, M. A., & Harris, P. L. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development*, *76*, 1261–1277. doi:10.1111/j.1467-8624.2005.00849.x

Lenth, R. (2020). emmeans: Estimated marginal means, aka least-squares means (R Package Version 1.5.2-1) [Computer software]. Retrieved from https://CRAN.R-project.org/package=emmeans

Leslie, S.-J. (2008). Generics: Cognition and acquisition. *Philosophical Review*, *117*, 1–47. doi:10.1215/00318108-2007-023

Leslie, S.-J. (2017). The original sin of cognition: Fear, prejudice, and generalization. *Journal of Philosophy*, *114*, 393–421. doi:10.5840/jphil2017114828

Liberman, Z., Woodward, A. L., & Kinzler, K. D. (2017). The origins of social categorization. *Trends in Cognitive Sciences*, *21*, 556–568. doi:10.1016/j.tics.2017.04.004

Long J. A. (2019). interactions: Comprehensive, user-friendly toolkit for probing interactions (R Package Version 1.1.0) [Computer software]. Retrieved from https://cran.r-project.org/package=interactions

Müller, M. (2017). here: A simpler way to find your files (R Package Version 0.1) [Computer software]. Retrieved from https://CRAN.R-project.org/package=here

Noveck, I. A. (2001). When children are more logical than adults: Experimental investigation of scalar implicatures. *Cognition*, *78*, 165–188. doi:10.1016/s0010-0277(00)00114-1

Papafragou, A., & Musolino, J. (2003). Scalar implicatures: Experiments at the semantics–pragmatics interface. *Cognition*, *86*, 253–282. doi:10.1016/s0010-0277(02)00179-8

R Core Team. (2019). *R: A language and environment for statistical computing* (Version 3.6.1) [Computer software]. Vienna, Austria: R Foundation for Statistical Computing.

Rhodes, M., & Leslie, S.-J. (2018). Children's media communicates essentialist views of science and scientists. *PsyArXiv*. doi:10.31234/osf.io/s6chb

Rhodes, M., Leslie, S.-J., Bianchi, L., & Chalik, L. (2018). The role of generic language in the early development of social categorization. *Child Development*, *89*, 148–155. doi:10.1111/cdev.12714

Rhodes, M., Leslie, S.-J., Saunders, K., Dunham, Y., & Cimpian, A. (2018). How does social essentialism affect the development of inter-group relations? *Developmental Science*, *21*(1), Article e12509. doi:10.1111/desc.12509

Rhodes, M., Leslie, S.-J., & Tworek, C. M. (2012). Cultural transmission of social essentialism. *Proceedings of the National Academy of Sciences, USA*, *109*, 13526–13531. doi:10.1073/pnas.1208951109

Shafto, P., Goodman, N. D., & Frank, M. C. (2012). Learning from others: The consequences of psychological reasoning for human learning. *Perspectives on Psychological Science*, *7*, 341–351. doi:10.1177/1745691612448481

Stiller, A. J., Goodman, N. D., & Frank, M. C. (2015). Ad-hoc implicature in preschool children. *Language Learning and Development*, *11*, 176–190. doi:10.1080/15475441.2014.927328

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., . . . Yutani, H. (2019). Welcome to the tidyverse. *The Journal of Open Source Software*, *4*(43), Article 1686. doi:10.21105/joss.01686

Wilke, C. O. (2020). cowplot: Streamlined plot theme and plot annotations for 'ggplot2' (R Package Version 1.1.0) [Computer software]. Retrieved from https://CRAN.R-project.org/package=cowplot