COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 46 (2022) e13223

ISSN: 1551-6709 online
DOI: 10.1111/cogs.13223

Speaking of Kinds: How Correcting Generic Statements can Shape Children's Concepts

Emily Foster-Hanson,[a] Sarah-Jane Leslie,[b] Marjorie Rhodes[c]

[a]*Department of Psychology, Princeton University*
[b]*Department of Philosophy, Princeton University*
[c]*Department of Psychology, New York University*

Received 30 March 2022; received in revised form 23 September 2022; accepted 18 November 2022

Abstract

Generic language (e.g., "tigers have stripes") leads children to assume that the referenced category (e.g., *tigers*) is inductively informative and provides a causal explanation for the behavior of individual members. In two preregistered studies with 4- to 7-year-old children ($N = 497$), we considered the mechanisms underlying these effects by testing how correcting generics might affect the development of these beliefs about novel social and animal kinds (Study 1) and about gender (Study 2). Correcting generics by narrowing their scope to a single individual limited beliefs that the referenced categories could explain what their members would be like while broadening the scope to a superordinate category (Study 2) uniquely limited endorsement of gender norms. Across both studies, correcting generics did *not* alter beliefs about feature heritability and had mixed effects on inductive inferences, suggesting that additional mechanisms (e.g., causal reasoning about shared features) contribute to the development of full-blown essentialist beliefs. These results help illuminate the mechanisms by which generics lead children to view categories as having rich inductive and causal potential; in particular, they suggest that children interpret generics as signals that speakers in their community view the referenced categories as meaningful kinds that support generalization. The findings also point the way to concrete suggestions for how adults can effectively correct problematic generics (e.g., gender stereotypes) that children may hear in daily life.

Keywords: Cognitive development; Generic language; Essentialism; Social categorization; Animal categories; Gender; Explanation; Normative judgments

---

Funding information: Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health under Award Numbers F31HD093431 and R01HD087672; National Science Foundation grant BCS 1729540.

Correspondence should be sent to, Emily Foster-Hanson, Department of Psychology, Princeton University, 221 Peretsman-Scully Hall, Princeton, NJ 08540, USA. Email: emily.fosterhanson@princeton.edu

1. Introduction

Generic language (e.g., "tigers have stripes," "girls like pink") is common in child-directed speech (Brandone & Gelman, 2009; Gelman et al., 2004, 2008; Goldin-Meadow et al., 2005; Pappas & Gelman, 1998) and powerfully influences children's beliefs about categories (Cimpian & Markman, 2009, 2011; Gelman et al., 2010; Leshin et al., 2021; Rhodes et al., 2012). When children hear a category described with generic language, they become more likely to hold essentialist beliefs about the category—expecting that its members are highly similar to one another and share a wide range of properties due to a common intrinsic, underlying nature (Gelman, 2003; Gelman et al., 2010; Leshin et al., 2021; Rhodes et al., 2012). The mechanisms that undergird the influence of generic language on essentialist beliefs, however, are not yet well understood. Here, we consider one process by which generics might elicit essentialist beliefs in children—by signaling to children that speakers in their community view the category as an informative, meaningful kind that supports generalizations. To do so, we consider the implications of this proposed process for how to respond to generics to limit the spread of inaccurate and problematic essentialist beliefs about the social world.

Generic language robustly relates to essentialist beliefs about categories. In experimental contexts, when children are introduced to new categories via a series of generic (e.g., "Zarpies have striped hair") rather than specific (e.g., "This Zarpie has striped hair") descriptions, they develop stronger beliefs that category features are broadly shared across members (e.g., that Zarpies have other things in common, beyond striped hair; Benitez et al., 2022; Gelman et al., 2010; Leshin et al., 2021; Rhodes et al., 2012), that category features are inflexible (e.g., a Zarpie *cannot* have nonstriped hair; Benitez et al., 2022; Leshin et al., 2021; Roberts et al., 2017), and that category features are intrinsic and innate (e.g., inherited from Zarpie parents rather than caused by social mechanisms; Gelman et al., 2010; Leshin et al., 2021; Rhodes et al., 2012). In naturalistic settings, adults produce more generics to describe categories for which they themselves hold essentialist beliefs (e.g., for animal species and some social categories vs. for artifact categories or ad hoc groupings), and children correspondingly develop more essentialist beliefs about the categories and domains that they frequently hear described with generics (Brandone & Gelman, 2009; Gelman et al., 2008; Goldin-Meadow et al., 2005; Pappas & Gelman, 1998; Segall et al., 2015). On an individual level, the proportion of parents' category references that are in generic form correlates with the strength of their children's essentialist beliefs about those categories (Gelman et al., 2004; Meyer & Gelman, 2016; for similar findings concerning the relationship between teachers' language and their students' beliefs, see Rhodes et al., 2020; Wang et al., 2022).

Generic statements and the beliefs they induce about categories can sometimes facilitate children's learning (Diesendruck, 2003; Gelman, 2003; Gelman & Coley, 1990; Keil, 1989), but they also have broader implications in the social realm. Although generic descriptions of social categories (e.g., "boys love to wrestle," or "girls like dolls") often sound innocuous, they can contribute to the development of problematic essentialist beliefs about social categories. For example, children infer from such generic claims that the mentioned features should *not* be found in members of the other category (e.g., that boys *do not* like dolls, Moty & Rhodes, 2021), that category members *should* have the mentioned features (Roberts et al., 2017), and that something intrinsic about being a category member causes them to have shared features, and therefore, that they likely have other (as of yet discovered) features in common as well (Cimpian & Markman, 2009, 2011; Gelman et al., 2010; Leshin et al., 2021; Rhodes et al., 2012). In this way, generics set the stage for the acquisition of social stereotypes and other problematic inter-group beliefs and behaviors (Haslam et al., 2002; Leslie, 2017; Prentice & Miller, 2007). Because generic descriptions of categories are common, understanding the various mechanisms through which they shape concepts can illuminate how an adult might respond to them to reduce these negative consequences. The goal of the present studies is therefore to shed light on a novel mechanism by which generics shape children's essentialist beliefs and, in the case of social categories, experimentally test strategies for how speakers might respond to generic claims to limit the development of problematic beliefs.

As an example, imagine a parent who is shopping with their daughter for a birthday present for a friend, and a store clerk suggests a new doll, providing the explanation that, "girls love dolls." In this scenario, perhaps the parent knows that the friend in question does indeed like dolls but wants to follow-up on the store clerk's claim with their daughter anyway. The parent might wish to do so because he knows the daughter herself does not like dolls—or perhaps her brother *does* like dolls—and the parent does not want either of them to feel bad about their preferences. The parent might also want to challenge the general idea that what children like to play with is determined by their gender.

But what type of response is likely to be effective? The answer to this question depends on the processes by which generic language elicits and reinforces children's essentialist beliefs. One way that generics contribute to essentialism is by communicating that a feature (in this case, "liking dolls") is shared across members of a kind (*girls*). Children have a bias to explain category features in inherent terms (Cimpian & Salomon, 2014). Thus, when they hear statements like, "girls like dolls," "girls have long hair," "girls wear pink," and so on, they appeal to an inherent cause to explain each of these (Cimpian & Erickson, 2012; Cimpian & Markman, 2009, 2011), eventually leading to the over-hypothesis (e.g., Dewar & Xu, 2010; Gelman, 2003) that something intrinsic to girls gives rise to all their shared features. On this account, children infer the presence of a category essence as they search for explanations for their shared features. Given this process, corrections to claims about the relation between the category and the shared feature (e.g., "no, that's not right about girls" or perhaps "boys like dolls too") could prevent children from inferring that an intrinsic cause explains the shared feature (since they are now questioning if the feature is indeed shared across members of the kind) and thus prevent the initial generic from eliciting essentialist beliefs about gender.

We propose, however, that in addition to eliciting causal reasoning about shared features, generics *also* contribute to essentialism because the linguistic form itself signals to children that speakers in their community view the relevant category as a meaningful way to divide up the world. That is, although labeling categories (e.g., "This is a Zarpie!") is a basic mechanism for communicating which collections of things are categories (Waxman & Gelman, 2009; Waxman & Markow, 1995), category labels alone do not elicit essentialist beliefs (Rhodes et al., 2012). We propose that the choice to generalize features *across* a labeled category communicates the speaker's beliefs about the category's underlying structure, beyond the information conveyed by labels alone. If this is correct, then it is not enough to just question if girls *"like dolls,"* it is necessary to question if the category *girls* is a meaningful way to parse the world more generally. By age two, children recognize generics as referring to abstract kinds rather than to specific individuals (Gelman & Raman, 2003; Graham et al., 2011; Hollander, Gelman, & Star, 2002; Rhodes et al., 2017). Thus, they realize that the description "girls like dolls" refers to the abstract category of girls *in general* and not to a specific group of girls. Also, young children expect that knowledgeable speakers (i.e., adults) know the correct way to refer to categories (Csibra & Gergely, 2009; Danovich & Keil, 2004; Harris & Koenig, 2006; Jaswal, 2010, 2013; Jaswal et al., 2010; Landrum et al., 2015; Noyes & Keil, 2017; Putnam, 1975; VanderBorght & Jaswal, 2009; Vygotsky, 1978; Wilson & Keil, 1998). Therefore, in this example, a child could assume that the kind *girls* is objective and significant because the adults around them speak as though it is.

Another way of describing these processes builds on the linguistic literature on presupposition (e.g., Frege, 1892; Russell, 1905; Stalnaker, 1974; Strawson, 1950). For example, the statement "Zarpies have striped hair" may be understood to take for granted the existence of *Zarpies* as an informative kind that sensibly supports generalizations. In this way, generic statements could introduce the concept of *Zarpies*, as a meaningful kind, into the common ground (i.e., the body of information that the speaker and listener mutually assume to be true and not up for debate; Atlas, 2004; Heim, 1982; Karttunen, 1974; Karttunen & Peters, 1979; Soames, 1982; Stalnaker, 2002). Upon hearing such a generic statement, listeners might quietly accept this view of the category by adjusting their own beliefs (Lewis, 1979; Von Fintel, 2008; for a similar proposal about predicate nominals, see Ritchie, 2021). Thus, via this possible mechanism, listeners could conclude that the categories referenced by generics are meaningful kinds because an informed, reliable speaker sees them that way and has introduced this belief as common knowledge.

On this account, claims such as "girls like dolls" communicate more than information about who likes dolls; they also communicate that gender categories are targets for essentialist beliefs. Further facilitating these effects, by ages 4 and 5, children are beginning to actively construct essentialist theories of the social world—that is, they expect that some social categories reflect essential kinds and so are "on the look-out" for cues in their environment as to which categories have this structure. Within this context, generic language from a knowledgeable source serves as a strong cue that the referenced category is a good candidate.

From this perspective, any response to a generic that maintains the generic scope of reference—even if it challenges claims about the referenced features (such as, "no, that's not right about girls" or even "well, boys like dolls too")—are unlikely to limit the spread of essentialist beliefs. These responses challenge whether *particular features* are shared across members of the category, but because they maintain the generic form, they reinforce the idea that the referenced category is a meaningful way to divide up the world. Instead, on this account, it would be necessary to directly challenge the scope of reference by redirecting the conversation to focus on specific individuals (e.g., "well, *Mary* likes dolls").

The process that we test here—that generics elicit essentialist beliefs in part because the linguistic form, on its own, signals that speakers view the category as a meaningful and coherent kind—likely often operates in tandem with causal reasoning about the described features to account for how generics shape essentialist beliefs across the course of early childhood. But considering the particular role of the linguistic form makes unique predictions about when generics will elicit essentialism and how to counteract these effects. In particular, from this perspective, generics can elicit essentialist beliefs *even when* they do not provide any information about which features are shared across members of the kind. Thus, responses to generics that maintain a generic scope, even while contradicting other content (e.g., "that's not right about girls," "boys can do that too") are likely to still induce essentialist beliefs.

In the present studies, to investigate this proposed mechanism, as well as to reveal effective ways to respond to such generics, we compared the effectiveness of responses to generics that either challenge the generic scope of the claim or that maintain the scope but challenge the property information. In Study 1, we examined these questions in the context of novel categories, directly building on prior work on how generics elicit essentialist beliefs as children learn about new categories from other speakers. In Study 2, we examine these questions in a naturalistic context about gender categories—a common target of generic language in parent–child conversation and one of the first social categories that most children develop essentialist beliefs about.

2. Study 1

Study 1 examined how the *form* of generic statements might lead children to treat new groupings as meaningful kinds, beyond the *content* of the properties predicated by generic statements, by exposing children to generic sentences about novel categories that were later contradicted. Children heard generic statements in the form of a short puppet show; generic statements (e.g., "Zarpies have striped hair") were uttered by one puppet, and each was then immediately corrected by a second puppet. Puppet 2's corrections were of two different types, depending on the condition (which varied between-participants), with prosodic stress emphasizing the aspect of the original statement that was being corrected. For half of participants, these contradictions maintained the generic scope of the original sentence (uphold-scope condition, e.g., "No, no, no! *That's* not right about Zarpies"). The other half of the participants heard Puppet 2 correct the *scope* of Puppet 1's assertion to apply only to a single individual (limit-scope condition, e.g., "No, no, no! *This* Zarpie has striped hair.") For limit-scope condition corrections, the pattern of stress emphasizing "this" highlighted that only the scope of the original generic statement from Puppet 1 was being challenged, thereby undermining the category as a relevant locus of generalization.

We first conducted two small, exploratory studies with young children to develop our new methodological approach and to refine our hypotheses; these studies are fully reported in the Supplementary Online Materials on the Open Science Framework (https://osf.io/ukqvz). We then proceeded to conduct a large confirmatory test of our central hypothesis, which is reported here. The study design, hypotheses, and analyses were preregistered on OSF (https://osf.io/sz8me). We focused the current study on children aged 5–7 because by this age children regularly understand generic statements (Cimpian & Erickson, 2012; Cimpian & Markman, 2009, 2011; Gelman & Raman, 2003) and are engaged in the process of determining which categories mark meaningful kinds in their environments (Mandalaywala et al., 2019; Rhodes & Gelman, 2009).

We included three measures of children's category-related beliefs, including beliefs that category membership is (a) causally powerful (with a modified forced-choice explanation task), (b) intrinsic and stable (with the switched-at-birth task), and (c) inductively informative (with an inductive inference task). We also probed the robustness of our effects by examining the mechanisms by which generic language shapes category-based reasoning across domains by testing effects for both novel animal categories and novel social categories.

2.1. Method

2.1.1. Participants

A power analysis based on effects from previous related studies (e.g., Rhodes et al., 2012, Study 1b) indicated that we would need a total sample size of 107 participants per condition for 80% power; we planned to recruit a total of 220 participants to allow for possible exclusions.

Of our total recruited sample of 220 participants, 10 were dropped at the time of testing because they did not complete at least two out of three measures or because of technical malfunctions, parental interference, or significant interruptions during the study. Additionally, four participants were excluded from analysis after testing because they failed all memory check questions as specified in our preregistration plan. This left a total of 206 5- to 7-year-old children ($M_{\text{age}} = 6.30$; 93 boys; 113 girls). We recruited from a continuous age range of 5–7 years and include participant age (measured continuously) as a predictor variable in all models. Of our sample of children whose parents provided demographic information (17% chose not to report this information), 54% were White, 2% were Black, 16% were Asian, 1% were Native Hawaiian or Other Pacific Islander, 1% were Middle Eastern or North African, and 26% were more than one race; 22% of those reporting demographic information were Hispanic (of any race). All participants spoke English as their primary language. Children were recruited and tested at a science museum in a large northeastern city in the United States. For both Studies 1 and 2, parents provided written consent and children gave verbal assent, and children received a small prize for participating. All study procedures were approved by the Institutional Review Board of New York University (protocol IRB-FY2016-760, Conceptual development and social cognition).

2.1.2. Procedure

Participants learned about a novel category called "Zarpies," which, by random assignment, was either described as a kind of animal or a kind of person. Animal stimuli were those used by Gelman et al. (2010) and social kind stimuli were those used by Rhodes et al. (2012); see Fig. 1). Scripts and materials, including videos of the full procedure, are available on OSF, https://osf.io/ukqvz, and sample videos of participants completing the study are available to authorized researchers on Databrary, https://nyu.databrary.org/volume/836.

Introduction phase: The procedure was divided into three phases. In the first phase, children saw an illustrated image of four Zarpies and heard, "Here is a picture of some Zarpies." Each Zarpie was then indicated one at a time and a voice said, "This is a Zarpie." Children were asked to repeat the category name to ensure they had understood the introduction before

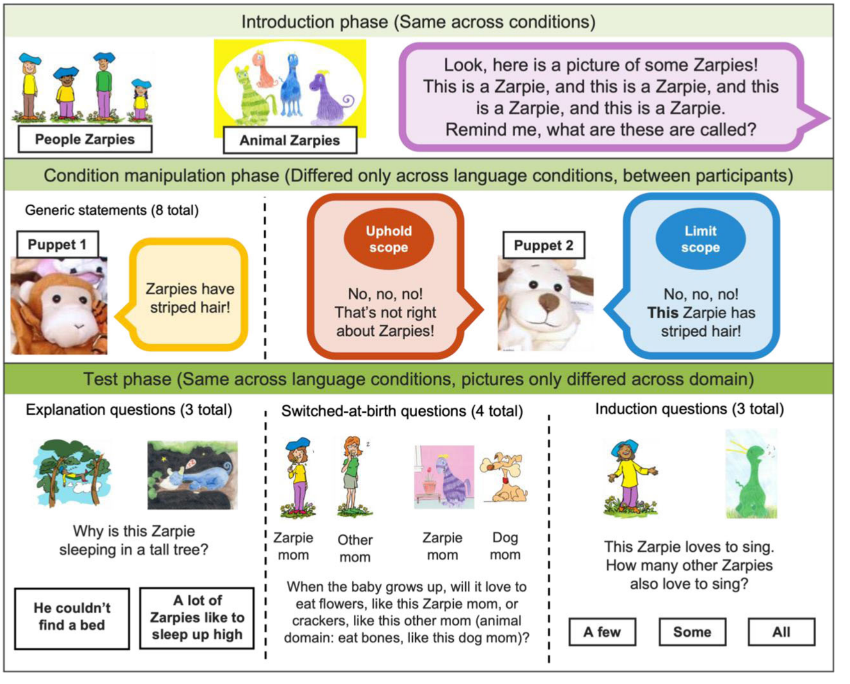

Fig. 1. Procedure, Study 1. Animal Zarpie stimuli were those used by Gelman et al. (2010) and people Zarpie stimuli were those used by Rhodes et al. (2012)). In the Introduction phase, children were introduced to the novel category called "Zarpies." In the condition manipulation phase, children heard generic statements from Puppet 1 and corrections (varied by condition) from Puppet 2. Finally, in the test phase, participants answered questions probing their beliefs about Zarpies.

moving on to the experimental phase. Children saw no other visual input about Zarpies until the test phase.

Condition manipulation phase: The second phase of the study contained the manipulation by condition in the form of a short puppet show with two animal puppets (a monkey and a dog). The puppet show was prerecorded and shown to the children as a video. First, the experimenter introduced the two puppets as "Sally" (Puppet 1, the monkey) and "Jenny" (Puppet 2, the dog), and told participants that they would be playing a game where Puppet 1 looked inside a special book of pictures and told the child what she saw. But, they were told, Puppet 1 could not see very well and had forgotten her glasses, so Puppet 2 was going to help her out by telling her if she got it wrong. To ensure children understood the procedure, they then completed a warm-up round in which Puppet 1 incorrectly identified a picture of a red square, saying it was a green triangle and had to be corrected by Puppet 2. In this way, Puppet 2 was established as a more knowledgeable and reliable speaker in this context.

After the warm-up, the experimenter said that there were pictures of Zarpies in the book and that Puppet 1 was going to look at them one at a time and tell them what she saw. Puppet 1 then looked inside the book and uttered a series of eight generic statements about Zarpies (e.g., "Zarpies have striped hair") one at a time. The statements were modified from those used by Rhodes et al. (2012) and Gelman et al. (2010); each expressed a unique physical or behavioral property of Zarpies. After each generic statement uttered by Puppet 1, Puppet 2 also looked inside the book and disagreed with Puppet 1 in one of two ways, depending on the condition.

In the limit-scope condition, Puppet 2 corrected the statement by limiting its scope to only apply to one individual (e.g., “No, no, no! *This* Zarpie has striped hair”). In the uphold-scope condition, in contrast, Puppet 2 rejected the entire predicate, without directly challenging the scope (e.g., “No, no, no! That’s not right about Zarpies”). This correction did not supply any property that could be interpreted as generalizing to Zarpies, but it reinforced Puppet 1’s implicit claim that *Zarpies* constitutes a meaningful kind to generalize about. After each correction, the experimenter moved on to the next statement by Puppet 1, followed by another correction by Puppet 2, and so on.

Videos of the puppet show differed only by language condition and were identical across both animal and social domains. Domain (social kind, animal kind) was thus crossed in a 2 $\times$ 2 design with correction condition (limit-scope, uphold-scope). Participants heard a total of eight generic statements and their respective corrections. We checked participants’ memory for the statements after the fourth and eighth statements to ensure that they were paying attention and that they understood the condition manipulation. For the limit-scope condition, children were asked if Zarpies possessed the property expressed in Puppet 1’s generic statement, or if just one Zarpie possessed the property as per Puppet 2’s correction (e.g., “Do Zarpies have striped hair or does just that one Zarpie has striped hair?”). For the uphold-scope condition, children were asked if Zarpies possessed the property expressed by Puppet 1 (e.g., “Do Zarpies have striped hair?”). Regardless of their response, children were told the correct answer in the form of Puppet 2’s previous statement about the property (e.g., “Remember, *this* Zarpie has striped hair.”)

Test phase: After the puppet show, participants were asked a series of questions probing their beliefs about Zarpies.

To measure beliefs that the category *Zarpies* was causally powerful, participants were asked to choose between two explanations for a new property of one Zarpie (all test items referred to properties that were not mentioned at all in the preceding puppet video). For example, participants were asked, “Why is this Zarpie sleeping in a tall tree?” and selected one of two possible explanations for the property: a transient, extrinsic cause (e.g., “He couldn’t find a bed”; scored as 0), or an inherent, category-general cause (e.g., “A lot of Zarpies like to sleep up high”; scored as 1). The questions were similar to those used in previous work (Gelman et al., 2010; Rhodes et al., 2012; see also Benitez et al., 2022; Leshin et al., 2021), and the answer choices offered to children were based on common explanations generated by children in our preliminary studies. The order of the explanation choices was counterbalanced between participants. Each child completed three forced-choice explanation questions before moving on to the next measure (other items included asking why a Zarpie was chasing a shadow and why a Zarpie was scared of a ladybug). These questions were identical across domains.

To test beliefs that the category *Zarpies* was intrinsic and stable, we presented children with a switched-at-birth story about a baby that was born to a Zarpie mother but raised by a non-Zarpie mother. Children were asked two memory check questions (e.g., “Who took care of the baby?") to ensure they understood the measure and were paying attention. After the story, we asked participants (a) whether the baby was a Zarpie or not a Zarpie, and (b) whether the baby would share three properties of the Zarpie mother, or alternative properties of the adoptive mother. The story and measures for the social domain were adapted from Rhodes et al. (2012); for the animal domain, these were adapted from Gelman et al. (2010). These measures, therefore, differed slightly between domains. Specifically, in the social domain, the non-Zarpie mother was human (described as "this other mom"), but in the animal domain, the non-Zarpie mother was a dog (described as "this dog mom"). The properties attributed to the non-Zarpie mother in the social domain described common human properties, while in the animal domain they described typical properties of a dog. Participants were asked (1) whether the baby would flap its arms when it is happy like the Zarpie mom or clap its hands when it is happy, like the other mom (animal domain: wag its tail when it is happy like the dog mom), and (2) whether the baby would love to eat flowers like the Zarpie mom, or would love to eat crackers, like the other mom (animal domain: would love to eat bones like the dog mom). For this measure, predictions that the child would have the properties of the birth mother (the Zarpie) were considered kind-based responses because such responses indicate that children viewed category-linked properties as determined by birth and stable across development (scored 1; Gelman et al., 2010; Rhodes et al., 2012; Springer & Keil, 1989; Waxman et al., 2007).

To measure the extent to which participants viewed *Zarpies* as inductively informative, children completed three induction questions in which they were shown a novel property of a Zarpie (not mentioned in the video-based puppet show) and were asked how many other Zarpies shared the property. For example, participants were shown a picture of a Zarpie singing and were asked, "This Zarpie loves to sing. How many other Zarpies do you think also love to sing?" (other items included asking how many Zarpies could flip in the air and how many Zarpies hop over puddles). Responses were coded so that higher numbers indicate broader generalization ("a few," coded as 1; "some," coded as 2; and "all," coded as 3). These questions were identical across domains.

Participants completed the study on a laptop computer using Keynote with a trained experimenter. Participants' responses were recorded on paper by the experimenter at the time of testing. Testing sessions were videotaped and coded from video for reliability; intercoder agreement was 99%, with discrepancies resolved by the first author reviewing the video. According to our preregistration plan, we planned to look at results as a composite as well as separately. Preliminary analyses indicated that the patterns across measures were highly dissimilar ($\alpha = -0.024$); therefore, we report analyses of the measures separately rather than as a composite.

2.2. Results

2.2.1. . Forced-choice explanation task

We analyzed participants' responses on the forced-choice explanation task using a generalized linear mixed model with the lme4 package (Bates et al., 2015). We used the "glmer" function to run a binomial regression with the main and interactive effects of domain (people,

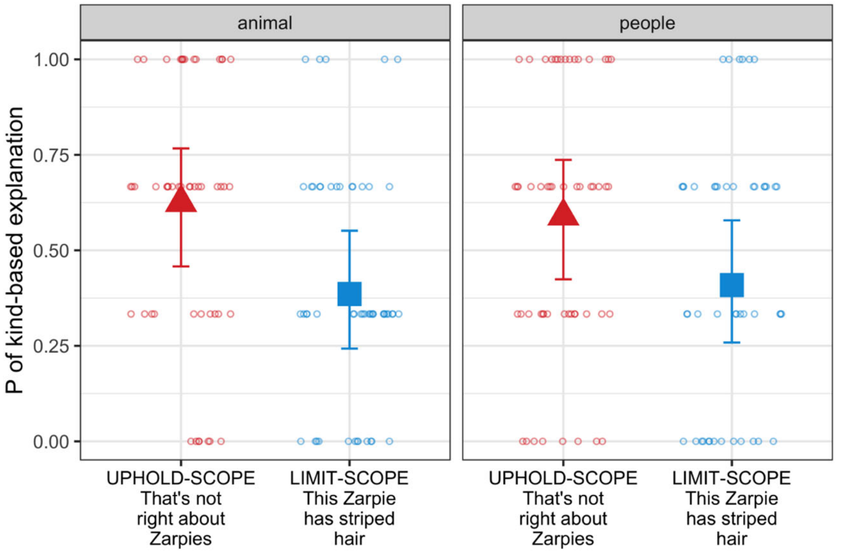

Fig. 2. Probability of endorsing kind-based explanations to explain properties of Zarpies in the forced-choice explanation task, by domain and condition, Study 1. Large shapes show group means with 95% confidence intervals; small circles show responses for each participant, averaged across the three explanation items.

animals) and language condition (limit-scope, uphold-scope) as predictors. Given the large age range of participants, we included participants' exact age in the model as a predictor as well. We also included random intercepts for each item and participant. Data and analysis code are available on OSF, https://osf.io/ukqvz. Means are reported as the probability of choosing kind-based responses with 95% confidence intervals. For all measures in this study, we report the results of likelihood ratio tests; this strategy is more conservative than the Wald chi-square tests described in our preregistration while allowing for consistency across measures and yielding the same pattern of results.

Participants were more likely to endorse kind-based explanations (e.g., "A lot of Zarpies like to sleep up high") in uphold-scope conditions ($M = 0.61$, 95% CI [0.46, 0.74]) than in limit-scope conditions ($M = 0.40$, 95% CI [0.26, 0.55]; main effect of language condition, X^2 (1) $= 14.59$, $p < .001$; Fig. 2). Relative to the limit-scope conditions, the odds of endorsing a kind-based explanation were 2.07 times as high (95% CI [1.11, 3.87]) in the uphold-scope conditions.

Children also endorsed more kind-based explanations with age (main effect of participants' exact age: X^2 (1) $= 8.48$, $p = .004$; there were no significant interactions with child age). Participants' explanation choices did not differ between the people and animal domains (X^2 (1) $= 0.02$, $p = .90$), nor did the domain interact with language condition (X^2 (1) $= 0.30$, $p =$ .58). When examined separately, the effect of language condition was similar and significant within each domain (people: X^2 (1) $= 4.57$, $p = .03$; animal: X^2 (1) $= 9.95$, $p = .002$), although the effect of age was significant only in the animal domain (X^2 (1) $= 9.98$, $p = .002$,

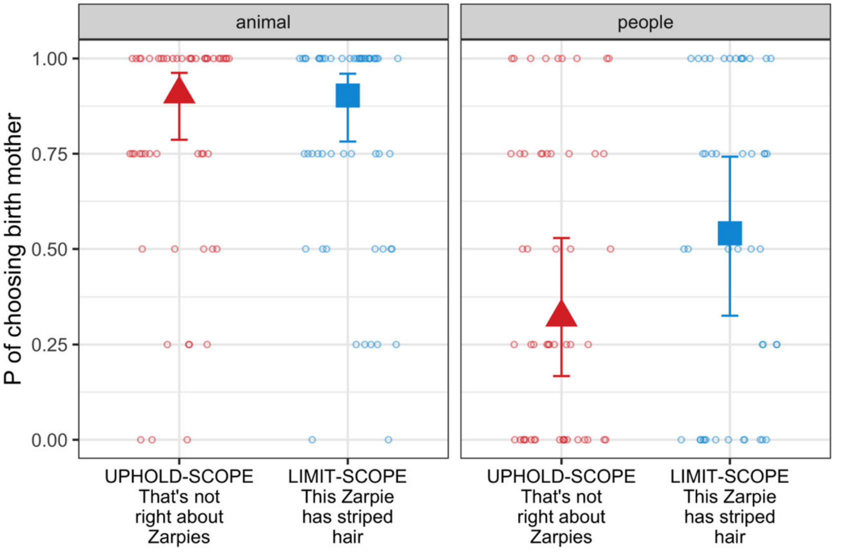

Fig. 3. Probability of responding that a Zarpie will be similar to its birth mother, rather than adoptive mother, in the switched-at-birth task, by domain and condition, Study 1. Large shapes show group means with 95% confidence intervals; small circles show responses for each participant, averaged across the four switched-at-birth items.

simple slope = .67, $p = .03$; people: X^2 (1) = 0.97, $p = .32$, simple slope = .14, $p = .60$). None of the means differed significantly from chance (all *p*s > .05).

2.2.2. *Switched-at-birth task*

We analyzed participants' responses in the switched-at-birth task as in the forced-choice explanation task. Participants were significantly more likely to predict that a Zarpie would display category membership and properties determined by its birth mother in the animal domain ($M = 0.90$, 95% CI [0.81, 0.95]) than for human Zarpies ($M = 0.43$, 95% CI [0.27, 0.59], the main effect of domain, X^2 (1) = 39.98, $p < .001$; Fig. 3). Relative to the people domain, the odds of giving a birth-mother response were 7.87 times as high (95% CI [2.46, 25.22]) in the animal domain. There were no main (X^2 (1) = 1.29, $p = .26$) or interactive effects (X^2 (1) = 1.41, $p = .24$) of language condition on this measure, nor were responses affected by participant age (X^2 (1) = 1.63, $p = .20$).

2.2.3. *Induction task*

We analyzed induction responses using an ordinal logistic regression model with the ordinal package (Christensen, 2019). Means are reported as average responses on the 1–3 scale with 95% confidence intervals.

In the people domain, participants generalized more broadly in the uphold-scope condition than in the limit-scope condition (X^2 (1) = 5.27, $p = .02$; Fig. 4). In the animal domain, however, generalizations did not differ by language condition (X^2 (1) = 0.01, $p = .90$). The overall condition x domain interaction was not significant, however, X^2 (1) = 1.92, $p = .17$.

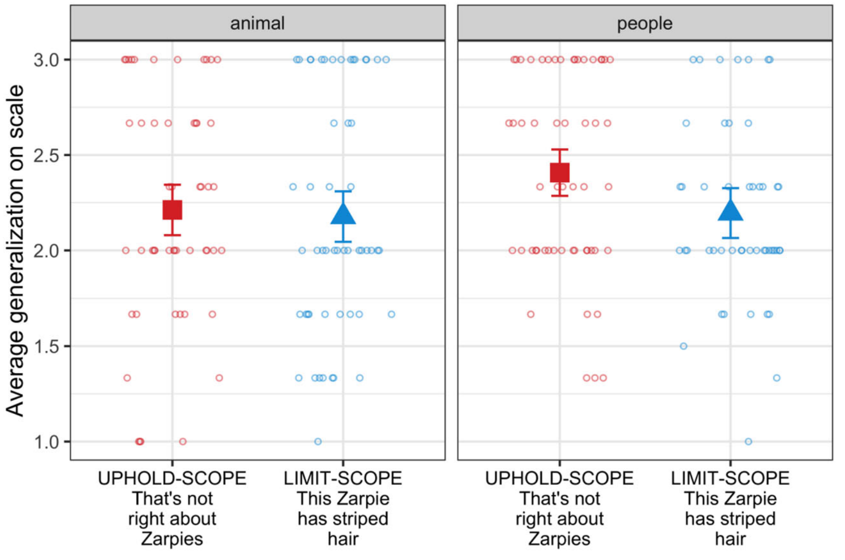

Fig. 4. Generalization of properties to other Zarpies in the induction task, by domain and condition, Study 1. Higher numbers indicate broader generalization ("A few" = 1; "Some" = 2; "All" = 3). Large shapes show group means with 95% confidence intervals; small circles show responses for each participant, averaged across the three induction items.

The effect of age was not significant in either domain (people: X^2 (1) = 1.03, p = .31; animals: X^2 (1) = 1.55, p = .21).

2.3. *Discussion*

In this study, responding to generics (e.g., "Zarpies have striped hair") with statements that challenged the scope of reference (e.g., "*this* Zarpie has striped hair") led children to endorse fewer intrinsic, category-based causes for new category features and to generalize new features of social categories more narrowly, relative to responses that maintained the scope but challenged the property description (e.g., "No, that's not right about Zarpies"). These findings are consistent with the possibility that one process by which generics elicit essentialist beliefs about categories is by signaling that speakers view the categories as a relevant way of dividing up the world, rather than only by communicating information about shared features because in this case, they did so even when they did not communicate information about which features are shared at all.

Although children endorsed more category-based, intrinsic causes in the uphold-scope than the limit-scope condition for both categories of animals and people, Puppet 2's response influenced beliefs about feature generalizability only for the categories of people. This difference across domains might reflect differences in children's baseline beliefs about these two domains (as noted in our preregistration). Young children already generally expect animal categories to have many features in common and therefore may be less reliant on linguistic cues as to whether a particular animal category has this structure. In contrast, children view only a small fraction of the social categories they encounter as inductively rich and reflecting deep homogeneity (Diesendruck et al., 2013; Kalish & Lawson, 2008; Rhodes & Gelman, 2009) and therefore maybe be more reliant on culture cues (in this case, generic language) to determine which categories these might be. Future research should directly test how generic language interacts with children's developing background beliefs about the structure of different types of categories (e.g., Kalish, 1998).

In the present study, language did not influence responses on the switched-at-birth task; instead, responses varied by domain, such that children viewed the properties of animal categories as more heritable and stable than those of human social categories. Indeed, children often view animal categories as more biologically determined than many human social categories (Rhodes & Gelman, 2009; Taylor et al., 2009), although the items for this particular measure differed by domain, making it somewhat difficult to interpret this main effect. Regardless, the lack of effect of the language manipulation on this measure may indicate that children's beliefs about category heritability may not be elicited by generic language unless the generics also communicate information about shared features (Gelman et al., 2010; Rhodes et al., 2012; see also Noyes & Keil, 2019). We will return to this point in the general discussion. Children's beliefs that the properties of animal Zarpies would be determined by birth (as measured in the switched-at-birth task) also increased with age, in line with the possibility that children's beliefs about biological heritability continue to develop across childhood (Springer & Keil, 1989; see also Menendez et al., 2022, for evidence that even adults are often unsure about the mechanisms of biological inheritance).

Study 1 was designed to test how generics lead children to treat kinds as meaningful by presenting them with categories that they would not view as particularly coherent and informative otherwise. Several features of the design make it striking that children began to develop essentialist beliefs about these categories at all in this context. Specifically, children saw very little visual input about the novel kinds in question (four diverse exemplars at the start of the experimental protocol, and no other visual input until the test phase), and they heard only eight generic statements about them. In this way, both the visual input and the number of generics was much less than in previous work (Leshin et al., 2021; Rhodes et al., 2012). Furthermore, participants received conflicting information from the two puppets. Generics only shape kind beliefs if they are uttered by a reliable speaker (Stock et al., 2009); similarly, listeners only accommodate new information conveyed implicitly by speakers "in good standing" (Von Fintel, 2008). Children were given a cover story and warm-up trials intended to lead them to view the second speaker as reliable, but nevertheless responses to generics that children might hear from trusted sources in everyday life might have stronger effects, since children are inclined to trust the testimony of adults unless they have a good reason not to (Hermansen et al., 2021; Jaswal & Neely, 2006; Wimmer et al., 1988). Finally, test items were designed to be either neutral or biased *against* kind-based responses. In all these ways, Study 1 may have underestimated the extent to which these processes can shape children's beliefs as they develop in daily life. Alternately, it is possible that these processes only operate as children are first learning about new categories (as in Study 1, when they were introduced to the novel category, "Zarpies") and are not strong enough to shape children's beliefs about categories that they may already be developing essentialist beliefs about. Study 2 began to address this question in the context of children's beliefs about gender.

3. Study 2

Study 2 tested the implications of the current proposed mechanism for how one might respond to generics to limit the spread of inaccurate and problematic essentialist beliefs about the social world. Parents regularly produce generic statements about gender categories when speaking to children (Gelman et al., 2004), and children hold strong beliefs about the coherence of gender categories by preschool age (Taylor, 1996; Taylor et al., 2009; Waxman, 2010). Parents' production of generic statements is correlated with their children's beliefs about gender in early childhood (Gelman et al., 2004), but no prior work has systematically measured the causal effect of generic statements on beliefs about gender categories throughout childhood. Beliefs about gender categories can also facilitate stereotyping and prejudice (Haslam et al., 2002; Leslie, 2017; Prentice & Miller, 2007), especially by giving rise to prescriptive judgments about what people can, should, and must do (Blakemore, 2003; Liben et al., 2002; Ruble et al., 2006; Rudman & Fairchild, 2004). Generic statements about novel social categories can elicit prescriptive expectations of conformity (Roberts et al., 2017) but, to our knowledge, no previous work has measured how generic statements about gender shape children's prescriptive gender norms. Study 2, therefore, tested the causal effect of generic statements on both descriptive and prescriptive beliefs about gender and whether contradicting the scope of generic statements about gender can reduce these beliefs in young children. Given the early emergence of essentialist beliefs about gender, we expanded our age range in Study 2 to include children aged 4 to 7 years.

To provide the most helpful practical advice for parents and caregivers who wish to limit the development of social stereotypes, Study 2 was presented in the naturalistic context of a conversation between a parent and child. Children regularly hear and utter generic statements in their daily lives (Gelman et al., 2004), meaning that even caregivers who avoid using generic statements about gender themselves will likely encounter these statements in the speech of their children or others in the child's social environment. For this reason, we also chose to have the child character utter the initial generic statement, followed by a response from the parent. This design had the added benefit of marking the corrective speaker (i.e., the adult) as more knowledgeable and reliable without the need for a warm-up phase.

3.1. Method

3.1.1. Participants

Based on the same power analysis used in Study 1, we intended to collect a total sample of 330 children (110 per condition) to allow for possible exclusions. We recruited children at a children's museum in a large urban northeastern city in the United States, as well as in local public pre-K centers and public schools. Due to the Covid-19 global pandemic, data collection at schools and museums was halted before we were able to recruit our total sample size. At that time, our recruited sample consisted of 313 participants.

Of our recruited sample, 17 participants were dropped at the time of testing because they did not complete at least three out of four measures or because of technical malfunctions, experimenter error, or parental interference (as specified in our preregistration). This left a total of 296 4- to 7-year-old children (*M* age = 5.98; 118 boys, 178 girls). Of our sample of children whose parents provided demographic information (35% chose not to report this information), 58% were White, 11% were Black, 18% were Asian, 1% were Middle Eastern or North African, and 11% were more than one race; 24% of participants were Hispanic or Latino (of any race). All participants spoke English as their primary language.

3.1.2. Procedure

Participants completed the study on a laptop computer using Powerpoint. Participants' responses were recorded on paper by a trained experimenter at the time of testing, and testing sessions were also videotaped and coded from video for reliability (sessions in school settings were audiotaped). Parents provided written consent and children gave verbal assent. Children received a small prize for participating. All study procedures were approved by the institutional review board of New York University. Scripts and materials, including videos of the full procedure, are available on OSF, https://osf.io/frua2, and a sample video of a participant completing the study is available to authorized researchers on Databrary, https://nyu.databrary.org/volume/836.

Condition manipulation phase: Participants heard generic statements about 16 gender-neutral properties (Liben et al., 2002), presented in the form of a cartoon conversation between a parent and child. Each participant was randomly assigned to hear these statements about either boys or girls. In the cartoon, each property trial was presented as a "picture" of something that happened at camp (e.g., a boy or girl playing checkers), about which the child character made a positive generic statement referencing the gender category (e.g., "boys/girls are good at playing checkers"). The parent character then immediately responded to the child character's statement before proceeding to the next property trial.

Children were randomly assigned to hear one of two types of responses from the parent character. Given that young children already view a wide range of properties as shared among members of gender categories (Taylor, 1996; Taylor et al., 2009; Waxman, 2010), we could not test the role of corrections that undermined shared features. Instead, in the uphold-scope, the parent responded by affirming the generic (e.g., "yes, *boys/girls* are good at playing checkers"). We chose to have the parent affirm the child's generic statements because we expect this to be a common parental response in naturalistic settings when speaking of nonstereotypical properties. This provided a baseline condition to test the effectiveness of the two correction strategies. In the limit-scope condition, the parent responded by narrowing the scope of the generic statement to a single individual (e.g., "Hmm, no, *Alfred/Susan* is good at playing checkers").

We initially preregistered Study 2's design, hypotheses, and analyses based on these two conditions (uphold-scope and limit-scope) of OSF (https://osf.io/5mf3j). However, after data collection had already begun, we realized that parents might also correct the scope of their children's generic statements in daily life by *broadening* the scope to the superordinate, nongendered category (e.g., "kids"), and that under certain circumstances, this type of

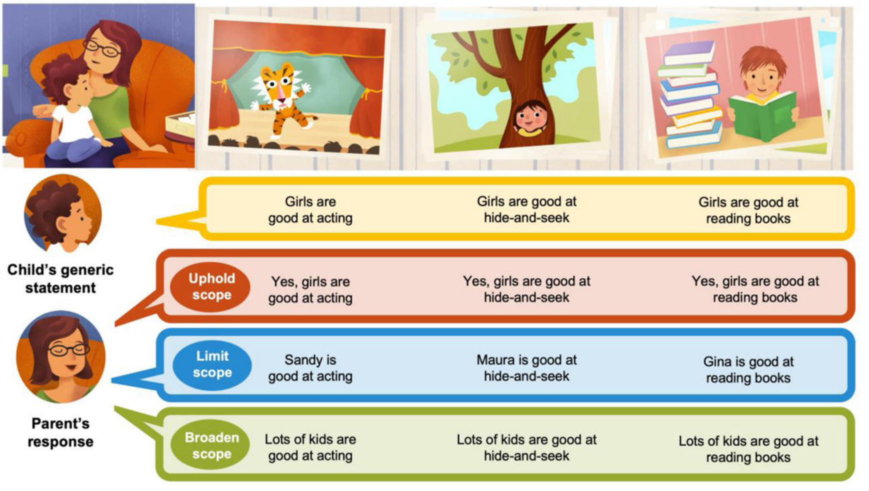

Fig. 5. Study 2 procedure, condition manipulation phase. Children watched a short cartoon of a parent–child conversation about "things that happened at camp." For each of 16 "pictures," the child character uttered a generic statement, and the parent character responded in one of three ways, depending on the condition.

correction might feel more natural for parents (like when correcting generic statements about stereotypes). To test the effect of this plausible type of correction, we preregistered a third condition, the broaden-scope condition, in which the parent responded by broadening the scope to the superordinate category (e.g., "Hmm, *lots of kids* are good at playing checkers"; https://osf.io/bk5e9). These corrections included the quantifier "lots" in order to be maximally naturalistic, since it is often untrue to say that kids as a category are good at various activities (i.e., it is not true that kids *in general* are good at singing), although we expected that children may nevertheless interpret these quantified statements as equivalent to generic statements about the superordinate category (Leslie & Gelman, 2012). These corrections were also more subtle than in the limit-scope condition because they did not involve any direct negation of the child character's generic statement, since a property that is true of a superordinate category (i.e., kids) would also be true of any subordinate categories (i.e., boys or girls). Thus, the gender category (boys, girls) was crossed in a 2 × 3 design with response condition (uphold-scope, limit-scope, broaden-scope). To make the language manipulation more naturalistic, the parent responded conversationally to the child's statement on the first four items (limit-scope and broaden-scope: "Who are you thinking of?"; uphold-scope: "When are you thinking of?") before correcting the statement (Fig. 5).

Test phase: After the cartoon, participants answered a series of questions probing their beliefs about the gender category they heard about during the condition manipulation phase. These questions varied by gender category for consistency (i.e., participants who heard generic statements and responses about boys were asked questions about boys). The questions were identical across parent response conditions (uphold-scope, limit-scope, broaden-scope). Our measured dependent variables consisted of three sets of questions probing the beliefs that gender categories are (a) causally powerful, measured using a modified forced-choice explanation task (four questions); (b) intrinsic and stable, measured using a switched-at-birth task; and (c) inductively informative, using an inductive inference task (three questions). We also included a measure of (d) flexibility, measured using a novel task in which a boy and a girl are pretending to be male and female fictional characters (four normative expectation questions and two prescriptive judgments of gender nonconformity, each followed by an open-ended explanation question coded for whether the child referred to gender norms).

To measure beliefs that the gender categories are causally powerful, children chose between two explanations for a novel property of the target gender (not mentioned in the previous condition manipulation phase), similar to Study 1. For example, participants heard, "There's a boy who has something called *osteoclasts* in his bones. Why do you think he has osteoclasts in his bones?" Children then selected one of two possible explanations for the property: a transient, extrinsic cause (e.g., "He fell down and got hurt," scored as 0) and an inherent, category-general cause (e.g., "A lot of boys have osteoclasts in their bones," scored as 1). The answer choices offered to children were based on common answers provided by children in our preliminary studies. The order of the explanation choices was counterbalanced between participants, and each child completed four forced-choice explanation questions before moving on to the next measure (other items included asking why a boy/girl had *thromboxane* in his/her brain, why a boy/girl was good at a sport called *leeming*, and why a boy/girl was good at a game called *toogit*.)

To test beliefs that gender categories are intrinsic and stable, we presented children with a switched-at-birth story, modified from that used by Taylor et al. (2009). Children were told a story about a baby who grew up only with members of the other gender category; for instance, they heard about a girl who was raised by her uncle and boy cousins and who went to a special school with only boys, so she only played with boys and never played with other girls. Children were then asked whether the baby would possess familiar stereotypical properties associated with the same or other gender (e.g., playing with a tea set or a toy truck; building with blocks or putting on makeup) and whether the baby would share novel properties with another child of the same gender or one of the other gender. These items were scored with properties of the same gender category coded as 1 and properties of the other gender category coded as 0 (Gelman et al., 2010; Rhodes et al., 2012; Springer & Keil, 1989; Waxman et al., 2007).

To measure the extent to which participants viewed gender categories as inductively informative, children completed three induction questions, based on those used by Waxman (2010). Participants saw color photographs of children selected from the California Facial Expressions dataset (LoBue, 2014; LoBue & Thrasher, 2015). For each item, children saw a photograph of a child of the target gender (the category they had heard statements about in the cartoon) and were told that he or she possessed a novel property (e.g., "This girl likes to eat a new snack called *naggles.*") They then saw a photograph of another child of the same gender and asked if she would also possess the property. Finally, they saw a photograph of a child of the other gender and asked whether he would possess the property as well. Responses were scored such that extending the property to the same gender but not the other gender was scored as 1, and all other patterns were scored as 0.

Last, we included an exploratory assessment of children's beliefs about the flexibility of gender categories, measured using a novel task in which a boy and a girl are pretending to be male and female superhero characters. Children were first asked which character each child should pretend to be (with same-gender responses coded as 1 and other gender responses coded as 0). They were then told that the children wanted to pretend to be other gender characters, and they were asked whether that was okay (coded as 0) or not okay (coded as 1). Children were then asked an open-ended explanation question about their judgment ("How come?"). Children's explanations were then coded for whether they included explicitly normative language, such as the modals *can*, *should*, or *must* (presence coded as 1, absence coded as 0). We also coded explanations for whether the child mentioned gender (presence = 1, absence = 0) and whether the explanation referred to personal choice (reference = 1, no reference = 0). All explanations were coded by a trained experimenter, with a randomly selected 35% coded by a second coder for reliability (initial agreement 93%; Cohen's Kappa = 0.86; disagreements resolved by the first author).

3.2. Results

3.2.1. Forced-choice explanation task

We analyzed participants' responses on the forced-choice explanation task as in Study 1, using the "glmer" function to specify a binomial distribution for the outcome variable and examining the effect of language condition (limit-scope, uphold-scope, broaden-scope). Given the large age range of participants, we included participants' exact age in all models as a predictor as well. As specified in our preregistration plan, we also included target gender category in these analyses, as well as random intercepts for items and participants. Data, analysis code, and study materials are available on OSF (https://osf.io/frua2). Means are reported as the probability of choosing gender-based explanations with 95% confidence intervals. For all measures in this study, we report the results of likelihood ratio tests; this strategy is more conservative than the Wald chi-square tests described in our preregistration while allowing for consistency across measures and yielding the same pattern of results.

In our primary analysis, there was a significant main effect of language condition (X^2 (2) = 7.00, p = .03): Overall, children in the limit-scope condition chose fewer gender-based explanations (e.g., "A lot of girls are born that way"; M = 0.26, 95% CI [0.17, 0.36]) than those in the uphold-scope condition (M = 0.34, 95% CI [0.24, 0.45]; pairwise contrast, p = .047) and directionally (but not significantly) fewer than the broaden-scope condition (M = 0.33, 95% CI [0.23, 0.44]; pairwise contrast, p = .12; Fig. 6). Children also endorsed fewer gender-based explanations with age (main effect of child age, X^2 (1) = 15.64, $p <$ .001) and fewer gender-based explanations to explain the features of girls (M = 0.27, 95% CI [0.18, 0.37]) than boys (M = 0.35, 95% CI [0.25, 0.46]; main effect of stimuli gender, X^2 (1) = 6.84, p = .009). There were no significant interactive effects; for additional exploratory analyses of patterns separated by child gender and property type, see the Online Supplementary Materials (https://osf.io/frua2).

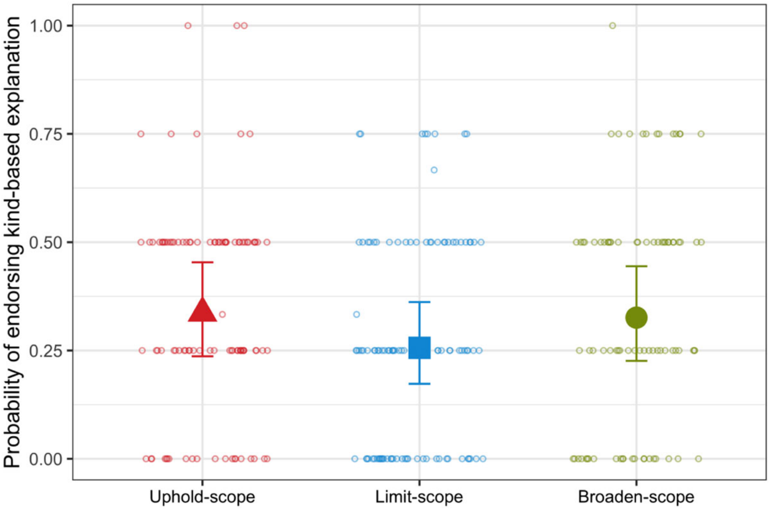

Fig. 6. Probability of endorsing kind-based explanations to explain properties of girls and boys in the forced-choice explanation task, by condition, Study 2. Shapes show group means, and error bars show 95% confidence intervals; small circles are responses for each participant averaged across the four explanation items.

3.2.2. *Switched-at-birth task*

We analyzed participants' responses in the switched-at-birth task as in the forced-choice explanation task; means are reported as the probability of responding that a child raised in an environment with only members of another gender category would nevertheless display their own gender category-typed features. Children chose fewer essentialist responses with age overall (X^2 (1) = 4.45, p = .04), but there were no main or interactive effects of language condition on this measure (Uphold-scope: M = 0.72, 95% CI [0.59, 0.83]; limit-scope: M = 0.77, 95% CI [0.65, 0.86]; broaden-scope: M = 0.66, 95% CI [0.51, 0.79]). For additional exploratory analyses of patterns separated by child gender and property type, see the SOM.

3.2.3. *Induction task*

We analyzed children's induction responses in the same manner as their explanation and switched-at-birth responses, with means as the probability of generalizing to another child of the same gender and not to another child of the other gender, with 95% confidence intervals. We found no significant main or interactive effects (uphold-scope: M = 0.18, 95% CI [0.13, 0.24]; limit-scope: M = 0.18, 95% CI [0.13, 0.25]; broaden-scope: M = 0.18, 95% CI [0.13, 0.25]; for details, see the SOM).

3.2.4. *Superhero task*

We first analyzed children's predictions about which superhero character the boy and girl should pretend to be, in the same manner as the other measures in this study; means are reported as the probability of saying that each child should pretend to be the same-gender character, with 95% confidence intervals. Children almost uniformly predicted that each child should pretend to be the same-gender character rather than the other-gender character (uphold-scope: $M = 0.99$, 95% CI [0.92, 1.00]; limit-scope: $M = 0.99$, 95% CI [0.93, 1.00]; broaden-scope: $M = 0.98$, 95% CI [0.88, 1.00]), and these expectations did not vary by any factor of interest.

We next analyzed judgments about whether it was okay for the children to pretend to be the other-gender characters, in the same manner as the other measures in this study; means are reported as the probability of saying that it was "not okay" for the children to pretend to be the other-gender characters with 95% confidence intervals. Children became more accepting of nonconformity with age (X^2 (1) = 17.22, $p < .001$), in line with previous work showing a decline across childhood in negative judgments of nonconformity to novel social groups (e.g., Foster-Hanson et al., 2021; Roberts et al., 2017). However, children's judgments also varied by language condition (X^2 (2) = 8.97, $p = .01$), with children in the broaden-scope condition viewing gender nonconformity as most acceptable. Children in the broaden-scope condition rarely said that nonconformity was "not okay" ($M = 0.14$, 95% CI [0.04, 0.37]), and they did so in this condition significantly less often than in the limit-scope condition ($M = 0.57$, 95% CI [0.35, 0.76]; pairwise contrast, $p = .02$) and directionally (but not significantly) less often than in the uphold-scope condition ($M = 0.27$, 95% CI [0.12, 0.51]; pairwise contrast, $p =$ .49).

The two-way interaction between condition and child age was not significant (X^2 (2) = 4.76, $p = .09$). However, given the main effect of age overall, and prior work showing age-related changes in judgments of nonconformity, we tested the effect of age separately in each condition. As shown in Fig. 7, responses in the broaden-scope condition did not vary by age (simple slope = -0.03, $p = .96$); in this condition, children of all ages reliably approved of nonconformity ($M = 0.14$, 95% CI [0.04, 0.37]; comparison to chance, $p < .001$). In contrast, there were effects of age in both the uphold-scope condition (simple slope = -1.60, $p = .007$) and limit-scope condition (simple slope = -1.33, $p = .008$). Younger children (aged 4–5) reliably disapproved of nonconformity in the limit-scope condition ($M = 0.57$, 95% CI [0.35, 0.76]; comparison to chance, $p = .009$) and were at chance in the uphold-scope condition ($M = 0.27$, 95% CI [0.12, 0.51]; comparison to chance, $p = .21$) and in both of these conditions, children became more approving with age (Fig. 7). In other words, children in both the limit-scope and uphold-scope conditions showed the general age-related change found in previous work (e.g., Foster-Hanson et al., 2021; Roberts et al., 2017) where younger children disapprove of category-related nonconformity and children become more approving with age. However, the broaden-scope condition appeared to disrupt this pattern by eliminating this effect of age, such that children in this condition approved of nonconformity even at the earlier ages.

Finally, we analyzed the open-ended explanations children gave for their normative judgments. We coded explanations for the use of explicit normative language about what people can, should, or must do, either to explain disapproval of nonconformity (e.g., "A boy has to be a boy and a girl has to be a girl") or approval (e.g., "Because they're allowed to be different people"). Normative language varied across conditions (X^2 (2) = 6.78, $p = .03$),

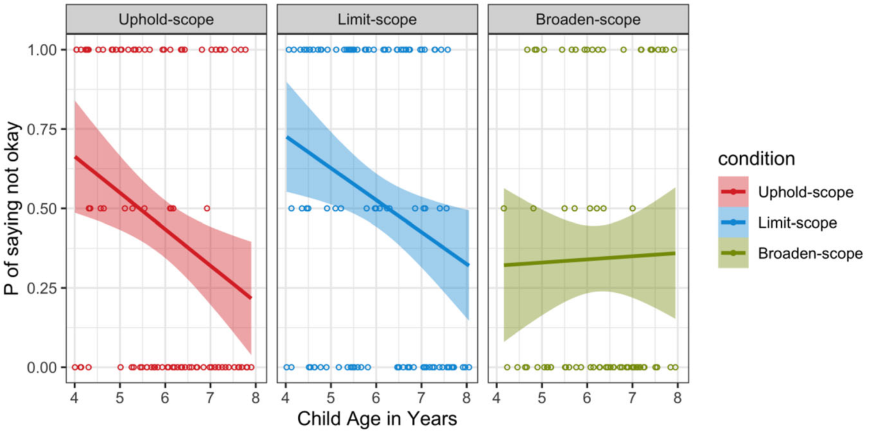

Fig. 7. Probability of saying that it was "not okay" for two children to pretend to be characters of the other gender, by correction condition and child age. Lines show group means with 95% confidence intervals; small circles show responses for each participant, averaged across the two trials. Children became more approving of gender nonconformity in both the limit-scope and uphold-scope conditions but approved of nonconformity consistently across ages in the broaden-scope condition.

with children using less normative language to explain their judgments of nonconformity in the broaden-scope condition ($M = 0.07$, 95% CI [0.02, 0.23]), compared to the limit-scope ($M = 0.24$, 95% CI [0.11, 0.46]; pairwise contrast, $p = .07$) and uphold-scope conditions ($M = 0.24$, 95% CI [0.11, 0.46]; pairwise contrast, p .08). This main effect was qualified by a two-way condition × target gender interaction (X^2 (1) = 7.75, $p = .02$), with children who heard about girls using the most normative language in the limit-scope condition, and children who heard about boys using the most normative language in the uphold-scope condition. Although variations in initial judgments across age and stimuli gender render this pattern somewhat difficult to interpret, in all cases normative language was more common in these two conditions than in the broaden-scope condition, providing further evidence that the broaden-scope condition led children to think of gender atypical behavior from a less normative perspective. There was also a significant two-way interaction between age and initial judgment (X^2 (1) = 13.23, $p < .001$), with older children using more normative language than younger children to justify approval (e.g., "They can do whatever they want") but no change across age in normative language to justify disapproval. There was also a two-way interaction between stimuli gender and initial judgment (X^2 (1) = 5.92, $p = .02$), with children who heard about boys using more normative language to explain disapproval and less to explain approval, compared to children who heard about girls, in line with previous evidence that gender nonconformity is judged more harshly for boys than girls (Connell, 1995; Koenig, 2018; Moss-Racusin, 2014; Sullivan et al., 2018). For additional analyses of coded explanations, see the SOM.

3.3. Discussion

Study 2 measured the causal effect of generic statements on beliefs about gender and tested two potential methods for correcting generic statements to limit gender essentialist beliefs in 4- to 7-year-old children. Children who heard a parent character limit the scope of a child's generic statements about gender categories (e.g., responding to "girls are good at painting" with "well, *Mary* is good at painting") subsequently endorsed fewer gender-based explanations for individual behavior than children who heard a parent character affirm the generic (e.g., "yes, girls are good at painting") or broaden the statement's scope to the superordinate category (e.g., "well, *lots of kids* are good at painting"). Broadening the scope of the claim to *kids* did, however, lead children to adopt less prescriptive beliefs about gender, in that they were more willing to say that gender nonconformity was okay and less likely to use normative language when explaining their thoughts about gender atypical behavior. Overall, these findings are striking because children in the current age range (4- to 7-year-olds) likely already held normative beliefs about gender and beliefs that gender categories can explain individual behavior (Cimpian & Markman, 2011; Taylor et al., 2009). These findings show that particular responses to generics can lead children to revise their existing beliefs, adopting more open and flexible views of how gender explains individual behavior than they would have otherwise.

In contrast to children's descriptive beliefs about gender (as captured by the explanation task), correcting generic statements by broadening their scope to the superordinate category was a uniquely successful strategy for lowering prescriptive expectations that gender constrains what people can, should, or must do. Children in the broaden-scope condition reliably approved of gender nonconformity across age, and they used less explicitly normative language to explain their nonconformity judgments. Previous work has found that generic statements can elicit prescriptive expectations of conformity among novel social categories (Roberts et al., 2017). However, to our knowledge, the current work is the first to show that generic statements also shape children's prescriptive beliefs about gender, and the first to test strategies to mitigate these potentially harmful effects.

One interpretation of this result is that children in Study 2 interpreted the generic statements themselves as communicating both descriptive regularities and prescriptive norms about how things ought to be (Haslanger, 2014; Leslie, 2015; McConnell-Ginet, 2011). People use generic statements (e.g., "boys don't cry") to communicate descriptive information (e.g., that it is uncommon for boys to cry—a descriptively false statement) as well prescriptive norms (e.g., that it is unacceptable or wrong for boys to cry). On this interpretation, children could have interpreted the parent character's corrections as not only correcting the descriptive regularity (e.g., something like, "It's not just common for girls to be good at riding bikes, it's common for kids in general to be good at ride bikes") but also as correcting the prescriptive norm (e.g., something like, "It's not just permissible and expected for girls to be good at riding bikes, it's permissible and expected for kids in general to be good at riding bikes"). What is most striking about the current result is that the parent character's corrections shaped participants' normative judgments about behaviors that they had not heard mentioned in the condition manipulation phase, meaning that they interpreted the corrections as communicating something important about how gender constrains what people ought to do *in general*. Also, this effect of condition was most apparent in the youngest children in our sample, while older children generally approved of nonconformity across conditions. This suggests that correcting statements about gender by broadening them to the superordinate category (or referring to superordinate categories rather than gendered categories to begin with) may be a fruitful tool for softening children's gender norms as they are actively developing.

In contrast to corrections that broadened the statement's scope, corrections that limited generic statements to a single individual (e.g., "well, *Mary* is good at painting") did not reduce children's normative judgments of gender nonconformity in Study 2. One possible interpretation is that while limit-scope corrections were effective at undermining the presupposition that the kind in question is an informative way of dividing up the social world, they do not speak to the prescriptive claims that generics also communicate. Prescriptive norms can be dissociated from descriptive regularities (or irregularities); for instance, the statement "boys don't cry" can be both descriptively false about what most boys are like while nevertheless communicating something true about how boys are supposed to be (Leslie, 2015). If children interpreted the child character's generic statements as both descriptive and prescriptive, they might have interpreted the parent character's limit-scope correction as agnostic about the prescriptive claims (e.g., something like, "This specific person is good at riding bikes, but we don't know what other category member are like or should be like"). That is, while broadening a normative generic statement could communicate that *everyone* is supposed to do the behavior in question, limiting it to a single individual instead questions whether *anyone* is supposed to do it.

4. General discussion

The present studies investigated the mechanisms by which generic language elicits essentialism by examining the comparative effectiveness of different ways of correcting generics. In Study 1, correcting generics by limiting their scope of reference (e.g., responding to "Zarpies have striped hair" with "No, *this* Zarpie has striped hair)" led children to think that the referenced categories had less explanatory power and were less inductively informative about the people (though not animals) who belonged to them, in contrast to correcting generics by retracting the property while leaving the scope intact ("No, that's not true about Zarpies"). These corrections did *not* differently influence children's beliefs about the extent to which the properties of Zarpies were inborn. In Study 2, responding to generic claims about gender (e.g., "Girls are good at painting") by limiting their scope (e.g., "Hmm, no, *Mary* is good at painting") again led children to think that gender had less explanatory power, while correcting the claims by broadening the scope of reference (e.g, "Hmm, *lots of kids* are good at painting") led them to apply fewer prescriptive norms to gender categories. As discussed below, these corrections did *not* influence children's beliefs about the extent to which gender-related properties were inborn, or their beliefs about whether new features would be broadly shared across gender categories.

By evaluating the effectiveness of these different responses, the current studies shed light on *how* generics shape children's concepts in the first place. In particular, these findings highlight how the generic *form* can shape children's beliefs, even in a manner separable from the content that generic claims convey. In Study 1, generics (e.g., "Zarpies have striped hair") led children to view novel social and biological kinds as having more explanatory power, even when the features they described were later falsified (e.g., "No, that's not right about Zarpies.") In this condition, children heard a series of generics that were later falsified by a more knowledgeable speaker, so at the end of the learning phase of the experiment, they had not learned about *any* properties that the novel category, "Zarpies," had in common with one another. Yet, they nevertheless thought of the novel category as having more explanatory power than children did in the condition where they heard corrections that affirmed the properties of individual Zarpies.

These findings, therefore, highlight how the generic form shapes children's beliefs, even when generics do not communicate information about shared features. We suggest they do so because when speakers refer to a kind using a generic statement, they communicate that they view the kind as meaningful, and they introduce this view as common knowledge (Atlas, 2004; Heim, 1982, 1983; Karttunen, 1974; Karttunen & Peters, 1979; Soames, 1982; Stalnaker, 1974, 2002). Children rely on the speakers in their community to teach them about the world and the kinds of things in it (Danovich & Keil, 2004; Jaswal, 2010, 2013; Jaswal et al., 2010; Noyes & Keil, 2017; Putnam, 1975; VanderBorght & Jaswal, 2009; Wilson & Keil, 1998), and they expect adults to be knowledgeable and reliable informants unless they have good reasons not to (Hermansen et al., 2021; Jaswal & Neely, 2006; Wimmer et al., 1988). Children are also "on the look-out" for cues about which kinds are meaningful as they use categories to learn and explain their experiences (Foster-Hanson & Rhodes, 2021; Gelman, 2003, 2004; Gelman & Roberts, 2017; Rhodes & Mandalaywala, 2017; Rhodes & Moty, 2020). When reliable speakers mark a kind as meaningful, children quietly adjust their own beliefs to match these cues (Lewis, 1979; Ritchie, 2021; Von Fintel, 2008). Thus, generics lead children to view kinds as meaningful because they communicate that others in their community view them as such.

In Study 1, using generics to refer to a kind shaped beliefs even when the generics did not communicate information about specific shared features. However, most generics that children hear in daily life *do* also communicate information about shared features. These sentences likely trigger multiple processes—beyond the current proposed mechanism—to have even more powerful consequences. For example, as described above, hearing multiple generic statements communicates information about property clusters, and learning that category members share a cluster of properties may lead children to generate the second-order inference, or over-hypothesis, that members of the category *in general* share many features (Dewar & Xu, 2010; Gelman, 2003). Generics also elicit explanatory processes that lead children to view kinds as meaningful, by marking properties as nonaccidental and central to category membership (Cimpian & Erickson, 2012; Cimpian & Markman, 2009, 2011; Haward et al., 2018, 2021; Prasada & Dillingham, 2006, 2009). Children's explanations for the causes of regularities among category members might also give rise to prescriptive inferences (e.g., reasoning that tigers have stripes because stripes are helpful for camouflage, which can license the belief that tigers *should* have stripes; Foster-Hanson & Lombrozo, 2022; see also Foster-Hanson & Rhodes, 2019, 2022). In daily life, these mechanisms likely operate in tandem to shape the development of the conceptual structure.

The consequences of falsified generics documented here, including eliciting the belief that categories are homogeneous and that categories are causally powerful, are consistent with essentialist views of category structure—the view that category membership stems from an intrinsic cause that is in place before birth, which causally constrains the development of individual category members—but also with other views of category structure. For example, children might expect categories to be homogeneous and to have explanatory power because of stable but extrinsic or structural factors that give rise to shared features (Vasilyeva et al., 2018). Thus, one possibility is that the generic form (on its own, free from content) signals to listeners that kinds are coherent and informative but not necessarily *why* this is the case (e.g., whether they hold this status because of a shared essence or more structural features). From this perspective, exposure to the linguistic form of generics leads listeners to adopt a view that a particular category is a possible essentialized kind but would not be sufficient to elicit full-blown essentialist beliefs. This possibility could explain why we did not find effects of language condition on the switched-at-birth measure in either Study 1 or 2—a measure that likely relies on stronger essentialist beliefs about a category. This account may also explain why we did not find any effects of language condition on children's inductive inferences based on gender in Study 2 since children in the current age range may already hold robust beliefs about which features are shared among gender categories (Taylor, 1996; Taylor et al., 2009; Waxman, 2010). On this account, when generics also communicate information about specific shared properties, however (as has been done in previous work—Gelman et al., 2010; Rhodes et al., 2012—and is common in daily life; Gelman et al., 2004), then additional processes are triggered that further facilitate the development of essentialist beliefs.

These category beliefs are an important element of conceptual development (Gelman & Coley, 1990), but they can also have pernicious consequences. For instance, expectations that animal kinds are highly homogeneous can be problematic for children's reasoning about within-species variability (Emmons & Kelemen, 2015; Shtulman & Schulz, 2008). Beliefs about the coherence of social categories, like race and gender, can have particularly stark consequences (Rhodes et al., 2017) by contributing to negative inter-group phenomena including social stereotyping (Bastian & Haslam, 2006; Diesendruck & Menahem, 2015; Gelman et al., 2007; Haslam et al., 2002; Hirschfield, 1996; Keller, 2005; Leslie, 2008, 2013; Mandalaywala et al., 2019; Pauker, Ambady, & Apfelbaum, 2010; Prentice & Miller, 2007; Williams & Eberhardt, 2008). Generics can also communicate prescriptive norms about how things *ought* to be (e.g., "boys don't cry"; Haslanger, 2014; Leslie, 2015; McConnell-Ginet, 2011), which can lead people to view categories as constraining what their members can, should, and must do, and young children may be particularly sensitive to prescriptive norms about categories. For example, in the current Study 2, many participants judged that it was "not okay" for children to even *pretend* to be other-gender characters in a game. In the context of these consequences, the mechanism documented here could be particularly significant: simply asserting a generic claim—even if the content is neutral or positive (e.g., "girls like pink," "girls are good at reading")—can alter how listeners think about the category. Thus, by leading people to expect category members to be highly similar to each other, generics lead listeners to adopt views of category structure that could make them more likely to adopt stereotypes, even if the generics do not express negative content themselves. Ritchie (2021) recently made a similar proposal about predicate nominals, arguing that statements like "Anna is a female" trigger a presupposition that underpins essentialist beliefs.

These processes might also be particularly powerful because they operate through rather subtle mechanisms. Indirect speech acts can be forceful tools of communication precisely because they introduce new information into the common ground without making direct claims (De Saussure, 2013; Holtgraves, 2002; Lee & Pinker, 2010). Introducing one's views via presupposition is akin to "acting as if" a desired state of affairs in the world were already the case (Thomason, 1990). Indeed, the process of accommodating this new information often goes unnoticed by the listener as long as the speaker is assumed to be a reliable informant (von Fintel, 2008). In this way, generic claims can perpetuate representations of social categories that give rise to social stereotypes without the listener (or perhaps even the speaker) realizing that they are participating in the transmission of these beliefs.

Further compounding the nature of this problem, the present findings suggest that simply negating generics is not sufficient to undermine their influence. For instance, if a child hears a generic statement (e.g., "girls hate math,"), Study 1's findings suggest that responding by rejecting the predicate (e.g., "*that's* not right about girls") might not be an effective strategy for lessening the statement's influence on children's beliefs about gender. The present studies also suggest two possible solutions to this problem, however. In particular, to lessen the possible negative consequences of the generic, one would need to directly challenge the generic scope of the sentence by limiting it to a specific person. For example, when a child hears (or utters themselves) a generic statement about a gender category, a parent might ask which particular person the child is referring to (e.g., "What person do you mean? Yes, Jimmy does like trucks"). To undermine the prescriptive effect of the generic, the parent might also broaden it to a superordinate category (e.g., "Lots of kids like trucks").

Given the ubiquity of generic statements about gender in children's environments (Gelman et al., 2004), and the potentially harmful consequences of gender beliefs on how children evaluate both others and themselves (Bian et al., 2017; Blakemore, 2003; Liben et al., 2002), future research is needed to build on the current findings by directly testing how and when correcting generic language might shape children's descriptive and prescriptive gender beliefs across development. The current results offer helpful practical advice for parents and caregivers who wish to limit the development of social stereotypes, but the exact nature of that advice, and the feasibility of changing parental speech, will need to be tested in future work. For example, it is unclear whether it would be feasible to teach parents or caregivers to limit their use of generics when referring to familiar social categories, like gender, about which they hold robust beliefs. Rather, to foster substantive change in children's development of beliefs that a person's gender predicts what they are like and constrains what they do, interventions may need to target parents' beliefs, as well as their linguistic manifestations, to alter the transmission of these beliefs across generations.

Acknowledgments

We thank Susan Gelman for sharing study materials and Wendy Wang, Dani Schuler, Rashel Reizin, Jennifer Ra, Daryl Ocampo, Anjali Mehta, Rachel Leshin, Samuel Krystal, Amanda Cardarelli, Kevin Bui, and Ericka Barroso for assistance with data collection. We are especially grateful to the families and staff at the American Museum of Natural History and the Children's Museum of Manhattan for their participation in this research. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the National Science Foundation.

Conflict of interest

The authors declare no conflict of interest.

Open Research Badges

This article has earned Open Data, Open Materials and Preregistered Research Design badges. Data, materials and the preregistered design and analysis plan are available at https://osf.io/ukqvz/, https://osf.io/frua2/, https://osf.io/sz8me/, https://osf.io/5mf3j/, https://osf.io/bk5e9.

References

Atlas, J. D. (2004). *Presupposition.* The Handbook of Pragmatics, Oxford, England: Blackwell, 29–52.

Bastian, B., & Haslam, M. (2006). Psychological essentialism and stereotype endorsement. *Journal of Experimental Social Psychology*, *42*, 228–235. https://doi.org/10.1016/j.jesp.2005.03.003

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Benitez, J., Leshin, R., & Rhodes, M. (2022). The Influence of Linguistic Form and Causal Explanations on the Development of Social Essentialism. Available at: https://doi.org/10.31234/osf.io/mwxej

Bian, L., Leslie, S. J., & Cimpian, A. (2017). Gender stereotypes about intellectual ability emerge early and influence children's interests. *Science*, *355*(323), 389–391. https://doi.org/10.1126/science.aah6524

Blakemore, J. E. O. (2003). Children's beliefs about violating gender norms: Boys shouldn't look like girls, and girls shouldn't act like boys. *Sex Roles*, *48*(9), 411–419. https://doi.org/10.1023/A:1023574427720

Brandone, A. C., & Gelman, S. A. (2009). Differences in preschoolers' and adults' use of generics about novel animals and artifacts: A window onto a conceptual divide. *Cognition*, *110*(1), 1–22. https://doi.org/10.1016/j.cognition.2008.08.005

Christensen, R. H. B. (2019). A Tutorial on fitting Cumulative Link Mixed Models with clmm2 from the ordinal Package. Tutorial for the R Package ordinal, Available at: https://cran.r-project.org/web/packages/ordinal/.

Cimpian, A., & Erickson, L. C. (2012). The effect of generic statements on children's causal attributions: Questions of mechanism. *Developmental Psychology*, *48*(1), 159–170. https://doi.org/10.1037/a0025274

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*(1), 14–25. https://doi.org/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*(2), 471–492. https://doi.org/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*(5), 461–480. https://doi.org/10.1017/S0140525x13002197

Connell, R. W. 1995. *Masculinities*. Cambridge, MA: Polity.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in cognitive sciences*, *13*(4), 148–153. https://doi.org/10.1016/j.tics.2009.01.005

Danovitch, J. H., & Keil, F. C. (2004). Should you ask a fisherman or a biologist?: Developmental shifts in ways of clustering knowledge. *Child Development*, *75*(3), 918–931. https://doi.org/10.1111/j.1467-8624.2004.00714.x

De Saussure, L. (2013). Background relevance. *Journal of Pragmatics*, *59*, 178–189. https://doi.org/10.1016/j.pragma.2013.08.009

Dewar, K. M., & Xu, F. (2010). Induction, overhypothesis, and the origin of abstract knowledge: Evidence from 9-month-old infants. *Psychological Science*, *21*(12), 1871–1877. https://doi.org/10.1177/0956797610388810

Diesendruck, G. (2003). Categories for names or names for categories? The interplay between domain-specific conceptual structure and language. *Language and Cognitive Processes*, *18*(5-6), 759–787. https://doi.org/10.1080/01690960344000116

Diesendruck, G., Goldfein-Elbaz, R., Rhodes, M., Gelman, S., & Neumark, N. (2013). Cross-Cultural Differences in Children's Beliefs About the Objectivity of Social Categories. *Child Development*, *84*(6), 1906–1917. https://doi.org/10.1111/cdev.12108

Diesendruck, G., & Menahem, R. (2015). Essentialism promotes children's inter-ethnic bias. *Frontiers in Psychology*, *6*, 1180. https://doi.org/10.3389/fpsyg.2015.01180

Emmons, N. A., & Kelemen, D. A. (2015). Young children's acceptance of within-species variation: Implications for essentialism and teaching evolution. *Journal of Experimental Child Psychology*, *139*, 148–160. https://doi.org/10.1016/j.jecp.2015.05.011

Foster-Hanson, E., & Lombrozo, T. (2022). How "is" shapes "ought" for folk-biological concepts. *Cognitive Psychology*. https://doi.org/10.1016/j.cogpsych.2022.101507

Foster-Hanson, E., & Rhodes, M. (2019). Is the most representative skunk the average or the stinkiest? Developmental changes in representations of biological categories. *Cognitive Psychology*, *110*, 1–15.

Foster-Hanson, E., & Rhodes, M. (2021). The psychology of natural kind terms. In S.T. Biggs, & H. Geirsson (Eds.), *The Routledge Handbook on Linguistic Reference*. London: Routledge.

Foster-Hanson, E., & Rhodes, M. (2022). Stereotypes as prototypes in children's gender concepts. *Developmental Science*, *139*, 101507. https://doi.org/10.1111/desc.13345

Foster-Hanson, E., Roberts, S. O., Gelman, S. A., & Rhodes, M. (2021). Categories convey prescriptive information across domains and development. *Journal of Experimental Child Psychology*, *212*, 105231. https://doi.org/10.1016/j.jecp.2021.105231

Frege, G. (1892). On sense and meaning. *Translations from the philosophical writings of Gottlob Frege*, *3*, 56–78.

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. Oxford, England: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195154061.001.0001

Gelman, S. A., & Coley, J. D. (1990). The importance of knowing a dodo is a bird: Categories and inferences in 2-year-old children. *Developmental Psychology*, *26*(5), 796–804. https://doi.org/10.1037/0012-1649.26.5.796

Gelman, S. A., Heyman, G. D., & Legare, C. H. (2007). Developmental changes in the coherence of essentialist beliefs about psychological characteristics. *Child Development*, *78*(3), 757–774. https://doi.org/10.1111/j.1467-8624.2007.01031.x

Gelman, S. A., & Raman, L. (2003). Preschool children use linguistic form class and pragmatic cues to interpret generics. *Child Development*, *24*, 308–325. https://doi.org/10.1111/1467-8624.00537

Gelman, S. A., & Roberts, S. O. (2017). How language shapes the cultural inheritance of categories. *Proceedings of the National Academy of Sciences*, *114*(30), 7900–7907. https://doi.org/10.1073/pnas.1621073114

Gelman, S. A., Taylor, M. G., & Nguyen, S. P. (2004). Mother-child conversations about gender: Understanding the acquisition of essentialist beliefs. *Monographs of the Society for Research in Child Development*, *69*(1), 93–118. https://doi.org/10.1111/j.0037-976X.2004.00279.x

Gelman, S. A., Ware, E. A., & Kleinberg, F. (2010). Effects of generic language on category content and structure. *Cognitive Psychology*, *61*(3), 273–301. https://doi.org/10.1016/j.cogpsych.2010.06.001

Gelman, S. A., Waxman, S. R., & Kleinberg, F. (2008). The role of representational status and item complexity in parent-child conversations about pictures and objects. *Cognitive Development*, *23*(2), 313–323. https://doi.org/10.1016/j.cogdev.2008.03.001

Goldin-Meadow, S., Gelman, S. A., & Mylander, C. (2005). Expressing generic concepts with and without a language model. *Cognition*, *96*(2), 109–126.https://doi.org/10.1016/j.cognition.2004.07.003

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. https://doi.org/10.1111/j.1467-8624.2010.01572.x

Haslam, N., Rothschild, L., & Ernst, D. (2002). Are essentialist beliefs associated with prejudice? *British Journal of Social Psychology*, *41*(1), 87–100. https://doi.org/10.1348/014466602165072

Haslanger, S. (2014). The normal, the natural and the good: Generics and ideology. *Politica & Società*, *3*(3), 365–392. https://doi.org/10.4476/78098

Harris, P. L., & Koenig, M. A. (2006). Trust in testimony: How children learn about science and religion. *Child Development*, *77*(3), 505–524. https://doi.org/10.1111/j.1467-8624.2006.00886.x

Haward, P., Carey, S., & Prasada, S. (2021). The formal structure of kind representations. *Cognitive Science*, *45*(10), Portico. https://doi.org/10.1111/cogs.13040

Haward, P., Wagner, L., Carey, S., & Prasada, S. (2018). The development of principled connections and kind representations. *Cognition*, *176*, 255–268.

Heim, I. (1982). *The semantics of definite and indefinite noun phrases*. University of Massachusetts, Amherst.

Heim, I. (1983). On the projection problem for presuppositions. In P. Portner & B. H. Partee (Eds.), *Formal Semantics–the Essential Readings* (pp. 249–260). Oxford, England: Blackwell Publishers

Hermansen, T. K., Ronfard, S., Harris, P. L., Pons, F., & Zambrana, I. M. (2021). Young children update their trust in an informant's claim when experience tells them otherwise. *Journal of Experimental Child Psychology*, *205*, 105063. https://doi.org/10.1016/j.jecp.2020.105063

Hirschfeld, L. A. (1996). *Race in the making: Cognition, culture, and the child's construction of human kinds*. Cambridge, MA: MIT Press.

Hollander, M. A., Gelman, S. A., & Star, J. (2002). Children's interpretation of generic noun phrases. *Developmental Psychology*, *38*(6), 883. https://doi.org/10.1037/0012-1649.38.6.883

Holtgraves T.M. (2002) *Language as social action: Social psychology and language use*. Mahwah, NJ: Erlbaum.

Jaswal, V. K. (2010). Believing what you're told: Young children's trust in unexpected testimony about the physical world. *Cognitive Psychology*, *61*(3), 248–272. https://doi.org/10.1016/j.cogpsych.2010.06.002

Jaswal, V.K. (2013). Biased to believe. In M.R. Banaji & S.A. Gelman (Eds.), *Navigating the social world: What infants, children, and other species can teach us* (pp. 241–244). New York: Oxford University Press.

Jaswal, V. K., Croft, A. C., Setia, A. R., & Cole, C. A. (2010). Young children have a specific, highly robust bias to trust testimony. *Psychological Science*, *21*(10), 1541–1547. https://doi.org/10.1177/0956797610383438

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*(9), 757–758. https://doi.org/10.1111/j.1467-9280.2006.01778.x

Kalish, C. (1998). Reasons and Causes: Children's Understanding of Conformity to Social Rules and Physical Laws. *Child Development*, *69*(3), 706–720. https://doi.org/10.1111/j.1467-8624.1998.tb06238.x

Kalish, C. W., & Lawson, C. A. (2008). Development of social category representations: Early appreciation of roles and deontic relations. *Child Development*, *79*(3), 577–593. https://doi.org/10.1111/j.1467-8624.2008.01144.x

Karttunen, l. (1974). Presupposition and linguistic context, *Theoretical Linguistics*, *1*(1-3), 181–194. https://doi.org/10.1515/thli.1974.1.1-3.181

Karttunen, L., & Peters, S. (1979). Conventional lmplicature. In C. -K. Oh & D. A. Dineen (Eds.), *Syntax and semantics 11: Presupposition* (pp. 1–56). New York: Brill.

Keil, F.C., 1989. *Concepts, Kinds and Development*. Bradford Books/MIT Press, Cambridge, MA.

Keller, J. (2005). In genes we trust: the biological component of psychological essentialism and its relationship to mechanisms of motivated social cognition. *Journal of Personality and Social Psychology*, *88*(4), 686. https://doi.org/10.1037/0022-3514.88.4.686

Koenig, A. M. (2018). Comparing prescriptive and descriptive gender stereotypes about children, adults, and the elderly. *Frontiers in Psychology*, *9*, 20180626. https://doi.org/10.3389/fpsyg.2018.01086

Landrum, A. R., Eaves Jr, B. S., & Shafto, P. (2015). Learning to trust and trusting to learn: A theoretical framework. *Trends in Cognitive Sciences*, *19*(3), 109–111.https://doi.org/10.1016/j.tics.2014.12.007

Lee, J.J., & Pinker, S. (2010). Rationales for indirect speech: The theory of the strategic speaker. *Psychological Review*, *117*, 785–807. https://doi.org/10.1037/a0019688

Leshin, R. A., Leslie, S. J., & Rhodes, M. (2021). Does it matter how we speak about social kinds? A large, preregistered, online experimental study of how language shapes the development of essentialist beliefs. *Child Development*, *92*(4), e531-e547. https://doi.org/10.1111/cdev.13527

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

Leslie, S. J. (2013). Essence and natural kinds: When science meets preschooler intuition. *Oxford Studies in Epistemology*, *4*, 108–166. https://doi.org/10.1093/acprof:oso/9780199672707.003.0005

Leslie, S. J. (2015). "Hillary Clinton is the only man in the Obama administration": Dual Character Concepts, Generics, and Gender. *Analytic Philosophy*, *56*(2), 111–141. https://doi.org/10.1111/phib.12063

Leslie, S. J. (2017). The original sin of cognition: Fear, prejudice, and generalization. *The Journal of Philosophy*, *114*(8), 393–421. https://doi.org/10.5840/jphil2017114828

Leslie, S. J., & Gelman, S. A. (2012). Quantified statements are recalled as generics: Evidence from preschool children and adults. *Cognitive Psychology*, *64*(3), 186–214.https://doi.org/10.1016/j.cogpsych.2011.12.001

Lewis, D.K. (1979). Scorekeeping in a language game. *Journal of Philosophical Logic*, *8* (1979), pp. 339–359

Liben, L. S., Bigler, R. S., Ruble, D. N., Martin, C. L., & Powlishta, K. K. (2002). The Developmental Course of Gender Differentiation: Conceptualizing, Measuring, and Evaluating Constructs and Pathways. *Monographs of the Society for Research in Child Development*, *67*(2), i–183. http://www.jstor.org/stable/3181530

LoBue, V., & Thrasher, C. (2015). The Child Affective Facial Expression (CAFE) set: validity and reliability from untrained adults. *Emotion Science*, *5*, 1532. https://doi.org/10.3389/fpsyg.2014.01532

LoBue, V. (2014). The Child Affective Facial Expression (CAFE) set. Databrary. Available at: https://doi.org/10.17910/B7301K

Mandalaywala, T. M., Ranger-Murdock, G., Amodio, D. M., & Rhodes, M. (2019). The nature and consequences of essentialist beliefs about race in early childhood. *Child Development*, *90*(4), e437-e453. https://doi.org/10.1111/cdev.13008

McConnell-Ginet, S. (2011). *Gender, sexuality, and meaning: Linguistic practice and politics*. New York: Oxford University Press.

Menendez, D., Mathiaparanam, O. N., Seitz, V., Liu, D., Donovan, A. M., Kalish, C. W., Alibali, M. W., & Rosengren, K. S. (2022). Like mother, like daughter: Adults' judgments about genetic inheritance. *Journal of Experimental Psychology: Applied*. https://doi.org/10.1037/xap0000436

Meyer, M., & Gelman, S. A. (2016). Gender essentialism in children and parents: Implications for the development of gender stereotyping and gender-typed preferences. *Sex Roles*, *75*(9), 409–421. https://doi.org/10.1007/s11199-016-0646-6

Moss-Racusin, C. A. (2014). Male backlash: Penalties for men who violate gender stereotypes. In R. J. Burke & D. A. Major (Eds.), *Gender in Organizations*. Cheltenham, UK: Edward Elgar Publishing.

Moty, K., & Rhodes, M. (2021). The unintended consequences of the things we say: What generic statements communicate to children about unmentioned categories. *Psychological Science*, *32*(2), 189–203. https://doi.org/10.1177/0956797620953132

Noyes, A., & Keil, F. C. (2017). Revising deference: Intuitive beliefs about category structure constrain expert deference. *Journal of Memory and Language*, *95*, 68–77. https://doi.org/10.1016/j.jml.2017.02.003

Noyes, A., & Keil, F. C. (2019). Generics designate kinds but not always essences. *Proceedings of the National Academy of Sciences*, *116*(41), 20354–20359. https://doi.org/10.1073/pnas.1900105116

Pappas, A., & Gelman, S. A. (1998). Generic noun phrases in mother-child conversations. *Journal of Child Language*, *25*, 19–33. https://doi.org/10.1017/S0305000997003292

Pauker, K.B., Ambady, N., & Apfelbaum, E.P. (2010). Race salience and essentialist thinking in racial stereotype development. *Child Development*, *81*, 1799–1833. https://doi.org/10.1111/j.1467-8624.2010.01511.x

Prasada, S., & Dillingham, E. M. (2006). Principled and statistical connections in common sense conception. *Cognition*, *99*(1), 73–112. https://doi.org/10.1016/J.COGNITION.2005.01.003

Prasada, S., & Dillingham, E. M. (2009). Representation of principled connections: A window onto the formal aspect of common sense conception. *Cognitive Science*, *33*(3), 401–448. https://doi.org/10.1111/j.1551-6709.2009.01018.x

Prentice, D. A., & Miller, D. T. (2007). Psychological essentialism of human categories. *Current Directions in Psychological Science*, *16*(4), 202–206. https://doi.org/10.1111/j.1467-8721.2007.00504.x

Putnam, H. 1975. *The Meaning of Meaning, Philosophical Papers, Vol. II: Mind, Language, and Reality*. Cambridge: Cambridge University Press.

Rhodes, M., Cardarelli, A., & Leslie, S. J. (2020). Asking young children to "do science" instead of "be scientists" increases science engagement in a randomized field experiment. *Proceedings of the National Academy of Sciences*, *117*(18), 9808–9814. https://doi.org/10.1073/pnas.1919646117

Rhodes, M., & Gelman, S. A. (2009). Five-year-olds' beliefs about the discreteness of category boundaries for animals and artifacts. *Psychonomic Bulletin & Review*, *16*(5), 920–924. https://doi.org/10.3758/PBR.16.5.920

Rhodes, M., Leslie, S.J., Bianchi, L., & Chalik, L. (2017). The role of generic language in the early development of social categorization. *Child Development*, *89*(1), 148–155. https://doi.org/10.1111/cdev.12714

Rhodes, M., Leslie, S., & Tworek, C. M. (2012). Cultural transmission of social essentialism. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, *109*(34), 13526–13531. https://doi.org/10.1073/pnas.1208951109

Rhodes, M., & Mandalaywala, T. M. (2017). The development and developmental consequences of social essentialism. *Wiley Interdisciplinary Reviews: Cognitive Science*, *8*(4), e1437. Portico. https://doi.org/10.1002/wcs.1437

Rhodes, M., & Moty, K. (2020). What is social essentialism and how does it develop? *Advances in Child Development and Behavior*, 1–30. https://doi.org/10.1016/bs.acdb.2020.05.001

Ritchie, K. (2021). Essentializing inferences. *Mind & Language*, *36*(4), 570–591. https://doi.org/10.1111/mila.12360

Roberts, S. O., Ho, A. K., & Gelman, S. A. (2017). Group presence, category labels, and generic statements influence children to treat descriptive group regularities as prescriptive. *Journal of Experimental Child Psychology*, *158*, 19–31. https://doi.org/10.1016/j.jecp.2016.11.013

Ruble, D. N., Martin, C. L., & Berenbaum, S. A. (2006). Gender Development. In N. Eisenberg, W. Damon, & R. M. Lerner (Eds.), *Handbook of child psychology: Social, emotional, and personality development* (pp. 858–932). New York: John Wiley & Sons, Inc.

Rudman, L. A., & Fairchild, K. (2004). Reactions to counterstereotypic behavior: the role of backlash in cultural stereotype maintenance. *Journal of Personality and Social Psychology*, *87*(2), 157. https://doi.org/10.1037/0022-3514.87.2.157

Russell, B. (1905). On denoting. *Mind*, *14*(56), 479–493.

Segall, G., Birnbaum, D., Deeb, I., & Diesendruck, G. (2015). The intergenerational transmission of ethnic essentialism: How parents talk counts the most. *Developmental Science*, *18*: 543–55. https://doi.org/10.1111/desc.12235

Shtulman, A., & Schulz, L. (2008). The relation between essentialist beliefs and evolutionary reasoning. *Cognitive Science*, *32*(6), 1049–1062. https://doi.org/10.1080/03640210801897864

Soames, S. (1982). How presuppositions are inherited: A solution to the projection problem. *Linguistic Inquiry*, *13*(3), 483–545. https://www.jstor.org/stable/4178288

Springer, K., & Keil, F. C. (1989). On the development of biologically specific beliefs: The case of inheritance. *Child Development*, *60*, 637–648. https://doi.org/10.2307/1130729

Stalnaker, R. (1974). Pragmatic presupposition. In M. Munitz & P. Unger (eds.), *Semantics and philosophy*. New York: New York University Press, 197–214. S. Davis (ed.), 471–81 (1991) (reprint).

Stalnaker, R. (2002). Common ground. *Linguistics and Philosophy*, *25*(5/6), 701–721. https://www.jstor.org/stable/25001871

Strawson, Peter F. (1950) On referring. *Mind 59*: 320–44. P.F. Strawson, 1–27 (1971) (reprint).

Stock, H.R., Graham, S. A., & Chambers, C.G. (2009). Generic language and speaker confidence guide preschoolers' inferences about novel animate kinds. *Developmental Psychology*, *45*, 884-888. https://doi.org/10.1037/a0015359

Sullivan, J., Moss-Racusin, C., Lopez, M., & Williams, K. (2018). Backlash against gender stereotype-violating preschool children. *PloS one*, *13*(4), e0195503. https://doi.org/10.1371/journal.pone.0195503

Taylor, M.G. (1996). The development of children's beliefs about social and biological aspects of gender differences. *Child Development*, *67*: 1555–1571. https://doi.org/10.1111/j.1467-8624.1996.tb01814.x

Taylor, M. G., Rhodes, M., & Gelman, S. A. (2009). Boys will be boys; cows will be cows: Children's essentialist reasoning about gender categories and animal species. *Child Development*, *80*(2), 461–481. https://doi.org/10.1111/j.1467-8624.2009.01272.x

Thomason, R. H. (1990). Accommodation, Meaning, and Implicature: Interdisciplinary Foundations for Pragmatics. In P. Cohen, J. Morgan, and M. Pollack (eds.), *Intentions in Communication*, Press/Bradford Books, Cambridge, MA, pp. 325–363.

VanderBorght, M., & Jaswal, V. K. (2009). Who knows best? Preschoolers sometimes prefer child informants over adult informants. *Infant and Child Development: An International Journal of Research and Practice*, *18*(1), 61–71. https://doi.org/10.1002/icd.591

Vasilyeva, N., Gopnik, A., & Lombrozo, T. (2018). The development of structural thinking about social categories. *Developmental Psychology*, *54*(9), 1735–1744 https://doi.org/10.1037/dev0000555

von Fintel, K. (2008). What is presupposition accommodation, again. *Philosophical Perspectives*, *22*, 137–170. https://www.jstor.org/stable/25177225

Vygotsky, L. S. (1978). *Mind in society: The development of higher mental process*. Cambridge, MA: Harvard University Press.

Wang, M. M., Cardarelli, A., Leslie, S.-J., & Rhodes, M. (2022). How children's media and teachers communicate exclusive and essentialist views of science and scientists. *Developmental Psychology*, *58*(8), 1455–1471. https://doi.org/10.1037/dev0001364

Waxman, S. (2010). Names will never hurt me? Naming and the development of racial and gender categories in preschool-aged children. *European Journal of Social Psychology*, *40*(4), 593–610. https://doi.org/10.1002/ejsp.732

Waxman, S., Medin, D., & Ross, N. (2007). Folkbiological reasoning from a cross-cultural developmental perspective: early essentialist notions are shaped by cultural beliefs. *Developmental Psychology*, *43*(2), 2. https://doi.org/10.1037/0012-1649.43.2.294

Waxman, S. R., & Gelman, S. A. (2009). Early word-learning entails reference, not merely associations. *Trends in Cognitive Sciences*, *13*(6), 258–263. https://doi.org/10.1016/j.tics.2009.03.006

Waxman, S. R., & Markow, D. B. (1995). Words as Invitations to Form Categories: Evidence from 12- to 13-Month-Old Infants. *Cognitive Psychology*, *29*(3), 257–302. https://doi.org/10.1006/cogp.1995.1016

Williams, M.J., & Eberhardt J.L. (2008) Biological conceptions of race and the motivation to cross racial boundaries. *Journal of Personality and Social Psychology*, *94*:1033–1047. https://doi.org/10.1037/0022-3514.94.6.1033

Wilson, R. A., & Keil, F. (1998). The shadows and shallows of explanation. *Minds and machines*, *8*(1), 137–159. https://doi.org/10.1023/A:1008259020140

Wimmer, H., Hogrefe, G. J., & Perner, J. (1988). Children's understanding of informational access as source of knowledge. *Child Development*, *59*, 386–396. https://doi.org/10.2307/1130318