

Does It Matter How We Speak About Social Kinds? A Large, Preregistered, Online Experimental Study of How Language Shapes the Development of Essentialist Beliefs

Rachel A. Leshin 
New York University

Sarah-Jane Leslie
Princeton University

Marjorie Rhodes 
New York University

A problematic way to think about social categories is to essentialize them—to treat particular differences between people as marking fundamentally distinct social kinds. From where do these beliefs arise? Language that expresses generic claims about categories elicits some aspects of essentialism, but the scope of these effects remains unclear. This study ($N = 204$, ages 4.5–8 years, 73% White; recruited predominantly from the United States and the United Kingdom to participate online in 2019) found that generic language increases two critical aspects of essentialist thought: Beliefs that (a) category-related properties arise from intrinsic causal mechanisms and (b) category boundaries are inflexible. These findings have implications for understanding the spread of essentialist beliefs across communities and the development of intergroup behavior.

When a child hears, “girls love pink,” “girls wear dresses,” “girls just don’t like math,” and so on, they learn (or mislearn) more than information about colors, clothing, and achievement—they learn that there is something about *being a girl* that fundamentally determines what a person is like. How profound are these effects of language? Children come to view some of the categories to which people belong—like gender—as marking fundamentally distinct kinds of people: people who are intrinsically different from one another, down to their essential core (Gelman, 2003; Haslam, Rothschild, & Ernst, 2000; Hirschfeld, 1996). Here we examined how these beliefs are fostered by commonplace patterns in language—particularly, generic descriptions of categories (saying that “girls love pink” instead of that a specific girl does so). To test this, we conducted a high-powered, preregistered experiment of children via a new webcam-based online laboratory, testing how language

shapes children’s beliefs about the meaning of otherwise arbitrary groups of people.

Languages Shapes Beliefs about Categories

A particularly problematic way to think about social categories is to *essentialize* them (Allport, 1954; Hirschfeld, 1996; Leslie, 2017; Rothbart & Taylor, 1992). For example, a person with an essentialist view of race thinks that people of the same race are fundamentally similar to one another and intrinsically different from people of a different race, that someone’s race determines (from even before birth) what they will be like, and that differences between racial groups are absolute, natural, and inevitable (Hirschfeld, 1996; Mandalaywala, Ranger-Murdock, Amodio, & Rhodes, 2019). These beliefs are problematic because they lead people to accentuate the differences between groups (Prentice & Miller, 2007) and assume that differences they see—including those related to status, wealth, achievement, or power, for example—reflect the essential capabilities or worth of the groups and their members, instead of more extrinsic or structural factors (Mandalaywala, Amodio, & Rhodes,

Research reported in this publication was supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development of the National Institutes of Health under Award Number R01HD087672. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Correspondence concerning this article should be addressed to Marjorie Rhodes, Department of Psychology, New York University, 6 Washington Place, rm. 301, New York, NY 10003. Electronic mail may be sent to marjorie.rhodes@nyu.edu.

© 2021 Society for Research in Child Development
All rights reserved. 0009-3920/2021/9204-0038
DOI: 10.1111/cdev.13527

2017). It is perhaps unsurprising, then, that social essentialism is associated with a range of problematic intergroup phenomena, including stereotyping (Yzerbyt, Corneille, & Estrada, 2001), prejudice (Haslam, Rothschild, & Ernst, 2002), decrements in memory for out-group faces (e.g., in the case of race; Gaither et al., 2014), perceptions of increased social distance between members of different groups (Diesendruck & Menahem, 2015), endorsement of a hierarchical status quo (Mandalaywala et al., 2017), reduced sharing of resources (Rhodes, Leslie, Saunders, Dunham, & Cimpian, 2017), and decreased motivation to interact with members of other groups (Pehrson, Brown, & Zagefka, 2009).

Children have the capacity to think in essentialist terms by ages three and four (Gelman, 2003); for instance, young children expect that a baby born to tiger parents will grow up to be a ferocious tiger, even if it is raised by a community of peaceful sheep (Gelman & Wellman, 1991; Waxman, Medin, & Ross, 2007). Such findings suggest that children view the behavior and potential of individual animals as determined by powerful, innate, and stable category “essences.” But while essentialist theories of the natural world reflect cognitive biases—not objective truths (Gelman & Rhodes, 2012; Leslie, 2013; Mayr, 1982, 1988)—they can nonetheless sometimes license correct and useful inferences (Gelman & Markman, 1986). For example, it is helpful to assume that a tiger will be ferocious, even if one hasn’t encountered that particular tiger before. Even within the biological domain, however, essentialized theories of categories can have problematic consequences, leading people to discount the natural variability present within biological categories and impeding scientific reasoning (Foster-Hanson & Rhodes, 2019; Gelman & Rhodes, 2012; Shtulman & Schulz, 2008).

Whereas the capacity for essentialist thought emerges early, children do not automatically apply an essentialist lens to the social world (for a review, see Rhodes & Mandalaywala, 2017). Taking race as an example, children develop essentialist beliefs about race slowly over the course of childhood (often between ages 5–10; Kinzler & Dautel, 2012; Mandalaywala et al., 2019; Rhodes & Gelman, 2009). Critically, whether and when children develop essentialist views of race depends on their own racial group membership (Kinzler & Dautel, 2012; Roberts & Gelman, 2016); the diversity of their neighborhood, schools, and communities (Mandalaywala et al., 2019; Rhodes & Gelman, 2009; see also Deeb, Segall, Birnbaum, Ben-Eliyahu, & Diesendruck, 2011; Smyth, Feeney, Eidson, &

Coley, 2017); and the political and social beliefs of their parents (Rhodes & Gelman, 2009; see also Segall, Birnbaum, Deeb, & Diesendruck, 2015). Which social groupings people view in essentialist terms also varies by culture and historical time—sometimes race, sometimes gender, sometimes religion, caste, ethnicity, geography, and so on (Hirschfeld, 1996; Rhodes & Mandalaywala, 2017). Although children possess early developing capacities for essentialist thought, cultural cues shape how (and when) they map these beliefs onto particular social divisions in their environment.

One potent cue that appears to guide the mapping of these beliefs is language. When children hear abundant generic claims about particular social groupings—claims that describe categories as abstract wholes, such as “*girls* don’t like math,” “*Italians* eat pasta,” or “*Hispanics* live in the Bronx”—they treat those ways of grouping people as more meaningful than they would otherwise (that is, compared to if they had heard only more limited descriptions of particular category members, such as, “this girl doesn’t like math,” “these Italians eat pasta,” or “this Hispanic family lives in the Bronx”; Foster-Hanson, Leslie, & Rhodes, 2019; Gelman, Ware, & Kleinberg, 2010; Segall et al., 2015). Indeed, parents’ use of generic language is associated with the development of their children’s essentialist beliefs. As one example, Segall et al. (2015) found that parents’ use of generic language in reference to ethnic categories (e.g., “Arabs”) predicted the extent to which children reported essentialist beliefs about those categories. In this case, the form of parents’ language in reference to social categories (i.e., whether it included generics or not) was a stronger predictor of children’s beliefs than the content that parents communicated about the categories themselves. Relatedly, Gelman, Taylor, and Nguyen (2004) found associations between mothers’ use of generic language about gender (e.g., “girls,” “boys”) and children’s own use of these terms. More generally, parents tend to use generics in domains in which they themselves hold essentialist beliefs about category structure (e.g., for animal species) more often than in domains in which they view categories as less coherent and stable (e.g., for artifacts, Gelman & Tardif, 1998; Gelman et al., 1998).

Although compelling, studies documenting correlations between parental language and children’s concepts cannot isolate the causal role of generic language in the development of social essentialism. To probe this causal mechanism, Rhodes, Leslie, and Tworek (2012) conducted a series of

experimental studies to see if generic language would elicit essentialist beliefs about a novel social category for which children would not have essentialist beliefs at the start (see also Gelman et al., 2010). In their study, Rhodes et al. (2012) introduced children (ages 4–5) to a new made-up group of people called “Zarpies.” The “Zarpies” were diverse in appearance (i.e., with respect to race, gender, and age), so if a child just looked at the group, it would appear as if members had been placed together arbitrarily, rather than as a product of some fundamental similarity or “essential core” shared among them. Therefore, baseline levels of essentialism were intended to be very low (close to zero) for this category in the absence of generic language. Children were introduced to Zarpies through storybooks that showed pictures of Zarpies one at a time; importantly, each picture was accompanied by a narrated description that either contained generics (e.g., “Look at this Zarpie! *Zarpies* sleep in tall trees.”), or didn’t (e.g., “Look at this Zarpie! *This Zarpie* sleeps in tall trees”). Many of the properties described in the storybook were designed to be unusual (e.g., “. . . draws stars on her knees,” “. . . really doesn’t like ice cream”)—thus further discouraging children from treating these properties as generalizable, innate, or stable in the absence of generic language (indeed, the tendency to treat these properties as individual and idiosyncratic at baseline was confirmed in various control conditions of the experiments).

Corroborating the correlations between generic language and essentialism found in studies of parental language (Segall et al., 2015), children who heard Zarpies described with generic language later viewed the category in more essentialist terms compared to children who heard specific language, thus demonstrating the power of generic language to elicit essentialist beliefs about social categories. While the effect sizes were modest, these studies show that even a small amount of generic language (presented in a 16-page storybook that was read to children 2–4 times, depending on the experiment) was sufficient to lead children to develop essentialist beliefs about a new perceptually diverse group when they would not otherwise (as confirmed in the control conditions). Furthermore, the effects of the language manipulation on children’s essentialist beliefs were documented several days after children heard the storybooks, indicating that generics did not just prime essentialist views in the moment, but rather, shaped how children learned to think about the new social group that they encountered (see also, Rhodes, Cardarelli, & Leslie, 2020). For

categories children encounter in their daily lives (e.g., gender categories, Gelman et al., 2004), the initially modest effects of generic language could likely compound over time. Furthermore, generics not only shape a listener’s beliefs; they also reflect a speaker’s thoughts. Indeed, parents who themselves hold essentialist beliefs about categories *produce* more generics in conversation with their children (Gelman & Roberts, 2017; Rhodes et al., 2012; Segall et al., 2015). Together, correlational studies of parental language (Gelman et al., 2004; Segall et al., 2015) and experimental studies (Gelman et al., 2010; Rhodes et al., 2012, 2017) identify generic language as a critical mechanism that shapes the development of essentialist beliefs.

How Deep Are These Effects?

And yet, a number of important questions remain about the scope and implications of these effects. Essentialism is comprised of a number of inter-related beliefs. For example, the particularly pernicious views of race described earlier include the beliefs that (a) people of the same race are fundamentally similar (i.e., that categories are *homogeneous*), (b) category-based similarities and differences are caused by *intrinsic* mechanisms, (c) someone’s race determines (from before birth) what they will be like (i.e., that differences are *innate* and *stable*), and (d) differences between groups are absolute, natural, and inevitable. It is possible to separate out these beliefs—for example, one might think that people of the same race are fundamentally similar to each other, but not believe that this is the *natural* state of the world or that group-based disparities arise from *intrinsic* causal mechanisms (Gelman, 2003; Noyes & Keil, 2019; Rhodes & Mandalaywala, 2017; Rhodes & Moty, 2020; Vasilyeva, Gopnik, & Lombrozo, 2018). Instead, these group patterns might be understood in terms of the structural and historical processes that led to them. Thus, the different components of people’s beliefs that fall under the umbrella of essentialism need not always relate to one another.

In previous research on generics and social essentialism, various measures of children’s concepts have been combined to yield composite scores of essentialism that collapse across multiple dimensions, without sufficient statistical power to examine the components separately. Thus, it is possible that generics shape some aspects of children’s beliefs (e.g., the idea that the category-related properties described by generics are generalizable, related to the idea that categories are *homogeneous*)

but not others (e.g., the beliefs that category-based regularities are intrinsic, innate, and inevitable, or that category boundaries are absolute and natural; Noyes & Keil, 2019, 2020; Saul, 2017; Vasilyeva et al., 2018).

If this is indeed the case, it would have important implications not only for how language shapes conceptual development, but also for some of the proposed social applications of this work. For example, consider the generic claim, “Girls get higher grades in English than math.” If one interprets this statement as just providing generalizable information (e.g., that more girls get higher grades in English), this might not itself be problematic. What really seems to matter is the causal attribution for this pattern. If one interprets this to mean that girls have a *natural* and *inherent* knack for English and deficiency in math, consistent with an essentialist framework, then this could lead to problematic stereotyping and discrimination (e.g., withholding from girls opportunities to succeed in math). But, if one interprets this statement as meaning that people have *previously (and unfairly) discouraged* girls in math, then this could instead lead to beneficial consequences—perhaps support for programs to rectify this inequality (Noyes & Keil, 2019; Rizzo, Elenbaas, & Vanderbilt, 2020; Vasilyeva et al., 2018). Similarly, if one takes the statement to describe absolute associations between categories and features (e.g., that a girl *could not* sometimes be better at math than English), this would be more problematic than if one interprets it only as a statistical association that does not constrain potential in this way. In other words, the problematic nature of generics seems to depend, at least somewhat, on which components of essentialism generics elicit—only beliefs that features are generalizable, or beliefs that category-related features are caused by *intrinsic mechanisms* and that these relations are absolute (and cannot be otherwise).

Thus, one goal of the present project was to identify whether or not generics elicit aspects of essentialist thought that go beyond the belief that category features are generalizable. Whereas some prior work has found that generics lead children to assume that the patterns described are explained by intrinsic causes (Cimpian & Markman, 2009, 2011; Foster-Hanson et al., 2019; Gelman et al., 2010; Rhodes et al., 2012, 2017), others have found effects only on whether people think the features described by generics are generalizable (Noyes & Keil, 2019, 2020, see also Vasilyeva et al., 2018), necessitating focused work on this question.

To address this issue, this study considered the effects of generic language on various distinct components of children’s beliefs as they learned about a new social group. In our exploration of these components, we opted not to include a measure of whether children think features are generalizable for the reasons indicated above. That is, the belief that features are generalizable might not always itself reflect essentialism (as people can think that category members share features due to extrinsic causes, as in the example above in which people could think that girls achieve less in math because of historic discrimination, not inherent potential). Second, thinking of features as generalizable might itself not be problematic; rather, as described above, the problematic consequences seem to depend on whether people endorse intrinsic mechanisms (e.g., thinking that girls are *inherently* worse at math) and constrain potential (e.g., thinking that a girl *cannot* be good at it). Instead, we used two common measures of whether children think that features described with generics reflect inherent causes—an explanation task (Cimpian & Markman, 2009, 2011; Rhodes et al., 2012) and a switched-at-birth task (which specifically tests if children think category-related features are inherited and stable; e.g., Hirschfeld, 1996; Taylor, Rhodes, & Gelman, 2009)—and one measure testing whether children think the relation between categories and their described features is absolute (and cannot be otherwise; see Rhodes et al., 2017; Taylor et al., 2009). In doing so, we sought to empirically test which aspects of essentialism are elicited by generics, a critical question both for conceptual development and for considerations of the social implications of generics for the development of children’s social cognition and behavior.

The Present Study

We tested these questions in a preregistered unmoderated remote experimental study (see Rhodes, Rizzo, et al., 2020; Scott, Chu, & Schulz, 2017; Scott & Schulz, 2017). The study was entirely self-administered; materials consisted of animations and narrations, and children completed the study remotely on their own time, without interacting with an experimenter. By automating the study in this way, we could achieve a high level of experimental control and facilitate replicability (for the benefits of unmoderated remote research for facilitating robust and reliable science, see Rhodes, Rizzo, et al., 2020). With this approach, we

recruited a larger, more geographically diverse sample in a much more efficient manner than could have been accomplished through traditional in-person testing; this increased our power to obtain precise estimates of the effect of generic language on each measure of essentialism, as well as to test for other covariates (e.g., child age). Also, though participation in the study necessitated access to a computer with WiFi and a webcam (a nonnegligible barrier to entry; see Lourenco & Tasimi, 2020), this online method allowed us to recruit a sample of children from neighborhoods with a high level of diversity with respect to population size, percent of racial-ethnic diversity, and median income. These metrics of diversity—though certainly not all-encompassing (in terms of participant race-ethnicity, for example, our sample was still quite homogeneous)—are particularly important within the context of our study given that social essentialism follows different developmental trajectories as a function of children’s local environments (e.g., Mandalaywala et al., 2019; Rhodes & Gelman, 2009). Thus, the ability to draw from a sample that did not come from a single geographic area had the potential to increase the generalizability of our findings (or to test for boundaries on these effects).

The original studies in Rhodes et al., 2012 focused on the ages at which social essentialist beliefs first emerge (4–5 years), but here we took advantage of the efficiency of our platform to extend the age-range upward to test if the implications of language for children’s beliefs remain consistent, increase, or decrease across age. While essentialist beliefs have sometimes been found to decrease across childhood (e.g., Astuti, Solomon, & Carey, 2004; Taylor, 1996; suggesting that linguistic mechanisms may less easily elicit these beliefs), children also often hold stronger essentialist beliefs across childhood about the particular categories for which adults in their community hold essentialist beliefs (Rhodes & Gelman, 2009), suggesting instead that these linguistic mechanisms could continue to shape children’s beliefs across age (Birnbaum, Deeb, Segall, Ben-Eliyahu, & Diesendruck, 2010; Deeb et al., 2011; Diesendruck, Goldfein-Elbaz, Rhodes, Gelman, & Neumark, 2015; Smyth et al., 2017).

Method

Participants

Participants were recruited to the online laboratory (the Princeton and NYU Discoveries in Action Lab, or *PANDA*; Rhodes, Rizzo, et al., 2020) via a

variety of methods, including community outreach efforts, social media posts, and advertisements on Facebook and parenting podcasts. We shared the link to the online laboratory broadly beginning in May 2019, and anyone who landed on the page with a child in the age-range of this study (4.5–8 years) was eligible to participate until November, when we closed the study. Additionally, a subset of our sample (34%) included children recruited through *Prolific*, a research platform based in the United Kingdom that posts studies for a registered pool of adult and child participants. Parents provided their child’s birthdate, gender, race, primary language, school level, and zip code.

As per our preregistration (<https://osf.io/3en4y>) we aimed to collect a sample of 200 children (approximately 100 per condition). This sample size was informed by a power analysis based on Rhodes et al., 2012, but then increased to allow us to detect changes across age, to identify the effects of a briefer language manipulation on essentialist beliefs (as here children heard the story containing the language manipulation only once, instead of 2–4 times as in previous research), to examine the various components of essentialism separately, and to explore possible regional variation in our findings. We specified in our preregistration that the age-range would include children ages 4.5–7.99, but because of a programming glitch, children ages 4.5–8.99 were able to participate. Because we had no theoretical reason to cut off the age-range, we decided to deviate from our preregistration and retain these slightly older children in the final sample. We also intentionally oversampled because we did not have a good way to estimate in advance the number of participants who might need to be excluded due to technical difficulties, parental interference, and so on.

With consent, children’s participation in the study was recorded via their computer’s webcam (and uploading a video at the end was required for inclusion in data analysis). The final sample included 204 children ($M_{\text{age}} = 6.36$, $SD_{\text{age}} = 1.17$; range = 4.50–8.87). An additional two children began the experiment and uploaded a video but completed < 50% of the test trials; as specified in our preregistration, these data were excluded from analyses.

These 204 participants were evenly balanced in terms of gender (50% female; 50% male) and included participants from a range of racial backgrounds (73.04% White, 5.39% Black, 3.43% Hispanic, 5.39% Asian, 12.75% multiracial). Due to the nature of our data collection, our sample reflected a

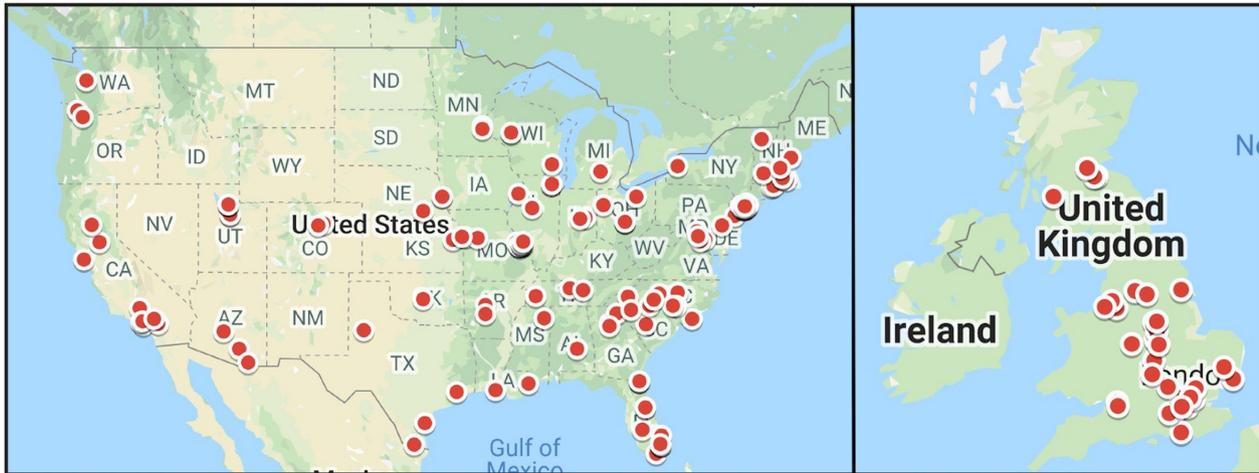


Figure 1. Regional diversity of participants across the United States and United Kingdom.

high degree of geographic diversity: 75.00% of children were recruited from the United States, 20.59% were recruited from the United Kingdom, 2.94% were recruited from Canada, and < 1% were recruited from Mexico, Australia, and New Zealand (see Figure 1). Even within the United States, our sample contains a high degree of geographic spread. Based on 2019 estimates from U.S. census data, the 153 U.S. participants who provided their zip code came from areas that varied substantially in population size ($M = 31,953$ residents, $SD = 18,125$; range = 775–89,453); proportion of Black, Asian, and non-White Hispanic residents ($M = 33.75\%$, $SD = 24.89\%$; range <1%–93%); and median family income ($M = \$97,916$, $SD = \$41,995$; range = \$36,933–\$215,405). Despite the neighborhood diversity reflected in these metrics, White children and children from higher income neighborhoods were overrepresented in the sample. For results broken down by country (i.e., United States vs. United Kingdom), as well as for more information about our neighborhood-level diversity metrics, see Supporting Information.

Procedure

The only three technical functionalities required for participation were (a) a webcam, (b) internet (with a sufficiently strong connection to upload a successful video at the end of the session), and (c) a Chrome or Firefox web browser. To begin the study, parents selected the study link and then clicked to allow the website to activate their webcams.

At the start of the study, parents provided verbal informed consent and children provided verbal

assent, following procedures approved by the IRB# FY2016-760. Then, parents and children were given instructions about how to navigate through the study—children were encouraged to click on answers themselves, or to point to their answers and have parents click for them if they were not able to. We encouraged parents not to provide any feedback to children during the study. At this point, children began the study, which consisted of a series of short videos and questions animated in Keynote, exported as movie files, and then programmed in Qualtrics. All stimuli files for the manipulation and test items may be viewed at <https://osf.io/acrwq/>.

As in the original study (Rhodes et al., 2012), we first introduced children to a novel social group called “Zarpies” via an illustrated storybook. The content of the storybook was identical to Rhodes et al. (2012): it contained 16 pages, each of which showcased a unique Zarpie displaying a particular physical or behavioral property (e.g., “likes to eat flowers” or “sleeps in tall trees”) that was narrated aloud (this time by a recorded narrator, instead of a live experimenter). The language used throughout the story varied based on the condition to which participants were randomly assigned: children in the *Generic* condition heard generic descriptions of Zarpies (e.g., “Look at this Zarpie! *Zarpies* like to eat flowers”), whereas children in the *Specific* condition heard language that referenced only the specific Zarpie on the page (e.g., “Look at this Zarpie! *This Zarpie* likes to eat flowers”). Children listened to the story aloud only once—a weaker manipulation than that used in the original study, where children heard the story 2–4 times, across a period

of several days. After listening to the story, children were asked a series of questions about Zarpies.

Measures

Our measures were comprised of three types of essentialism items (*explanation* items, *flexibility* items, and *inheritance* items), as well as two novel tasks probing intergroup cognition and behavior (*resource allocation* and *expectation of reciprocity* tasks). The design and composition of the essentialism items differed somewhat from previous work (Rhodes et al., 2012, 2017). As noted in the Introduction, we opted not to include measures of whether children think the properties described by generics are generalizable (previously tested with *induction* items) because these—on their own—are weak indicators of essentialist thinking. Instead, we focused on three measures that more clearly reflect essentialist thought (and that are also more likely to be problematic for social attitudes and behavior).

Essentialism Measures

Intrinsic causal mechanisms (explanation). The first measure of essentialist beliefs explored the extent to which participants viewed category membership as explaining the development of typical category properties—a central tenant of essentialist thought (Cimpian & Markman, 2011; Gelman, 2003; Taylor et al., 2009). In this task, children were asked a question about a behavior (“Why is this Zarpie hopping over a puddle?”) and provided two possible responses inspired by those commonly generated by children in previous research: one that invoked an intrinsic, category-related cause (“Because a lot of Zarpies think it’s fun to hop over puddles (scored as ‘1’)”), and another that invoked an extrinsic, individual reason (“Because she is practicing for a game of hopscotch (scored as ‘0’)”). This paradigm deviated from that used in Rhodes et al., 2012, wherein items were presented as open-ended questions and then coded for the presence of intrinsic and extrinsic explanations; here, we sought to preempt potential discomfort from speaking aloud to a computer screen by having children select, rather than generate, a response (for a similar approach, see Foster-Hanson et al., 2019).

Category boundaries as absolute (flexibility). Our second essentialism measure investigated children’s understanding of category boundaries as absolute and inflexible (Gelman, 2003; Rhodes & Gelman, 2009; Rhodes, Gelman, & Karuza, 2014; Roberts, Ho, Rhodes, & Gelman, 2017). Here we tested this

as the belief that a category member could only display properties associated with its category (and could never display properties not associated with it; similar to Taylor et al., 2009). For example, children were told, “Look at this Zarpie. This Zarpie sleeps in a tall tree. Do you think this Zarpie might sometimes sleep in a bed too (scored as ‘0’), or only in a tree (scored as ‘1’)?”.

Category features as innate and stable (inheritance). Our last measure of essentialism tested children’s belief in the innate potential of categories (i.e., the heritability of category membership and properties; Gelman, 2003) using a switched-at-birth paradigm (Gelman & Wellman, 1991; Hirschfeld, 1996; Waxman et al., 2007). As in Rhodes et al., 2012, we described a scenario in which a Zarpie baby was born to a Zarpie mom but raised by a non-Zarpie mom. Then, we asked, “Will this child like ice cream like the non-Zarpie mom (scored as ‘0’), or not like ice cream like the Zarpie mom (scored as ‘1’)?”. We chose to include one of the switched-at-birth questions from the original study (described above) and one directly asking if category membership is something one possesses “deep down” (i.e., “Is this [one] a Zarpie deep down (scored as ‘1’), or not a Zarpie deep down (scored as ‘0’)?”).

Intergroup Behavior Measures

Following the essentialism measures, children engaged in two additional tasks: one probing resource allocation decisions, and the other probing expectations of reciprocity. In both measures, children were introduced to a Zarpie and a non-Zarpie; then, children had to decide which target they (a) would offer a resource (e.g., a cookie) to, and (b) expect to receive a resource from. Each decision to give or expectation to receive a resource from a Zarpie was scored as “1”, and decisions to give or expectations to receive from a non-Zarpie were scored as “0”. Children made three decisions for each task.

The addition of these two measures was motivated by Rhodes et al. (2017), which found that inducing essentialist beliefs with generic language also led children to offer fewer resources to members of the newly essentialized kind. We reasoned that the inclusion of the resource allocation measure would allow us to conceptually replicate this effect, while the expectations of reciprocity measure may illuminate a potential mechanism underlying it (i.e., whereby children opt to give fewer resources to Zarpies because they expect to receive fewer in

return). Due to time constraints and differences in the testing platform, the details of our resource allocation task differed from the original study (Rhodes et al., 2017); in particular, children in our study were asked to allocate a series of resources to a Zarpie or a non-Zarpie one at a time, whereas in the original study, children were given a number of resources at once (e.g., 4 stickers) and asked *how many* to designate to a Zarpie.

Debriefing

After children had completed all measures, families were thanked for their time and participation and provided the opportunity to offer feedback on their study experience. As compensation for participation, families who completed the study received a \$10 Amazon giftcard (or, for participants recruited through Prolific, a digital payment of \$10). Overall, the study took families between 10 and 15 min to complete.

Video Coding

To ensure that the data collected in the online laboratory were valid—that is, that they captured real children participating, without undue influence from others—all videos were checked for the presence of a child and parent who, together, provided consent to participate. Additionally, 87% of all study videos were coded in Datavyu (Lingeman, Freeman, & Adolph, 2014) by a trained coder. Of these, 31% were double-coded by a second coder for reliability. Videos were coded for the presence of parent–child interactions (including pointing, verbal feedback, and gesturing) and other environmental distractions (e.g., pets making noise in the background), as well as the relevance of each of these instances. Critically, all audible coded interactions were evaluated for whether or not they interfered with the child’s response. Overall, the correlation of interference determinations per participant across primary and reliability coders was excellent ($r = .88$). Indeed, correlations between coders for trials coded as definitely or potentially containing an interference was high ($r = .75$), as were correlations between coders for interactions marked as not containing an interference ($r = .75$).

Notably, the number of interferences was small ($N = 21$ out of 2,492 trials completed by participants with coded videos; $< 1\%$ of total trials coded), and only one participant experienced interferences on more than 25% of trials. Because of the low incidence of interference, we chose to retain all

trials in our primary analyses—a deviation from our preregistration. For analyses with interference trials removed, as specified in our preregistration, see Supporting Information. Our full coding manual for Datavyu processing is available at <https://osf.io/acrwq/>. Additionally, 53.92% of families in our sample gave permission for their video to be shared with developmental researchers via Databrary (Rhodes, 2020).

Results

To examine the effects of language on each component of children’s essentialist beliefs, as well as their intergroup behaviors, we implemented a series of generalized linear mixed-effects models using the lme4 package in R (Bates, Mächler, Bolker, & Walker, 2015), with the experimental condition (Generic, Specific), age (entered as children’s exact age at the time of test, centered for analyses), and their interaction as fixed effects, and subject and item number as random effects. All analyses of essentialism and intergroup behavior were preregistered, but we made slight modifications to the models specified in our preregistration. First, since each of our measures contained multiple items, we opted to include the item number as an additional random effect. Second, due to the similarity of our effects across the region (i.e., United States and United Kingdom), we chose not to include children’s nationality as a covariate in our primary analyses. As noted in the Method, we instead feature analyses broken down by region in our Supporting Information. Data and analytic code for all analyses are available at <https://osf.io/acrwq/>.

Essentialism

Essentialism Composite

Replicating previous work (including Rhodes et al., 2012), we first considered the essentialism measures together, as a composite (we subsequently present analyses separately for each measure of essentialism). Consistent with findings from the original study, children endorsed more essentialist beliefs overall after hearing generic than specific language ($B = .48$, $SE = .13$, $z = 3.59$, $p < .001$; see Figure 2). Children in the Generic condition were 1.61 (Wald 95% CI [1.24, 2.12]) times as likely to endorse essentialist responses than children in the Specific condition. Children’s aggregate essentialism scores did not vary by age ($B = .12$, $SE = .08$,

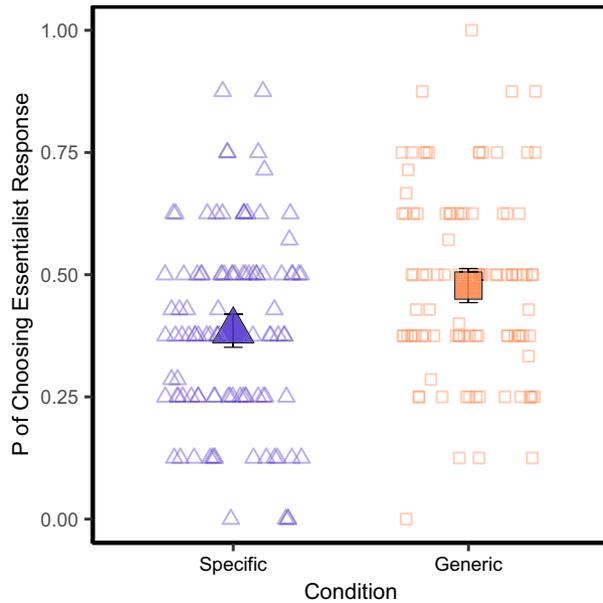


Figure 2. Participants' responses on the essentialism composite as a function of condition ($M_{\text{specific}} = .38$; $M_{\text{generic}} = .48$). Smaller shapes represent the proportion of essentialist responses given by each participant. Larger shapes represent the mean proportion of essentialist responses by condition; error bars represent 95% confidence intervals around these means.

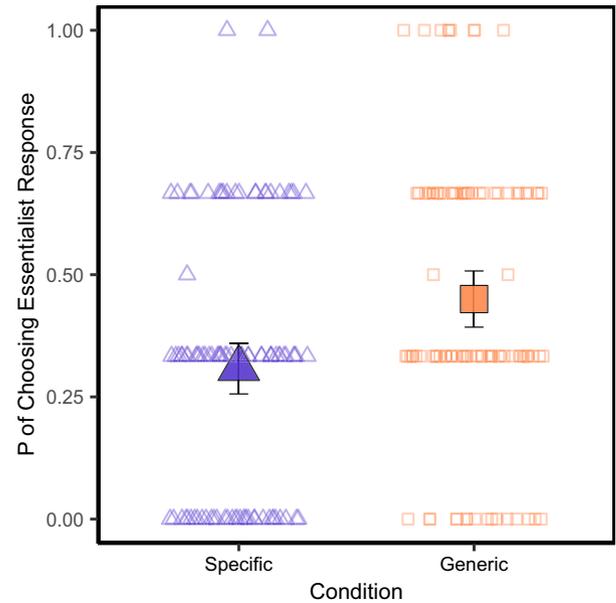


Figure 3. Participants' responses on the explanation items as a function of condition ($M_{\text{specific}} = .31$; $M_{\text{generic}} = .45$). Smaller shapes represent the proportion of essentialist responses given by each participant. Larger shapes represent the mean proportion of essentialist responses by condition; error bars represent 95% confidence intervals around these means.

$z = 1.46$, $p = .14$), nor as a function of an age by condition interaction ($B = .14$, $SE = .11$, $z = 1.19$, $p = .23$).

Intrinsic Causal Mechanisms (Explanation)

Children in the Generic condition endorsed more intrinsic explanations for Zarpie properties than those in the Specific condition ($B = .83$, $SE = .22$, $z = 3.77$, $p < .001$; OR = 2.34, CI [1.51, 3.77]; see Figure 3). Children also endorsed more essentialist responses with age ($B = .28$, $SE = .14$, $z = 2.03$, $p = .04$; see Figure 4), but there was no interaction between age and condition ($B = .22$, $SE = .19$, $z = 1.19$, $p = .23$).

Category Boundaries as Absolute (Flexibility)

Children in the Generic condition perceived the Zarpie category as more rigid and inflexible (meaning they were more likely to say that a Zarpie *could only* display features associated with being a Zarpie) than children in the Specific condition ($B = .54$, $SE = .20$, $z = 2.70$, $p = .007$; OR = 1.71, CI [1.16, 2.59]; see Figure 5). Effects on this measure did not vary by age ($B = .04$, $SE = .12$, $z = 0.36$, $p = .72$),

and there was no age by condition interaction ($B = -.08$, $SE = .17$, $z = -0.50$, $p = .62$).

Category Features as Innate and Stable (Inheritance)

Children's beliefs about heritability did not vary by condition ($B = -.09$, $SE = .26$, $z = -0.34$, $p = .73$), age ($B = .05$, $SE = .16$, $z = 0.33$, $p = .74$), or the interaction between the two ($B = .42$, $SE = .23$, $z = 1.84$, $p = .07$). For exploratory purposes (given previous work using this measure), we tested the slopes associated with age in each condition. Children's tendency to perceive the novel social category as heritable and stable increased with age in the Generic condition ($B = .45$, $SE = .16$, $z = 2.83$, $p = .005$), but not in the Specific condition ($B = .05$, $SE = .16$, $z = 0.32$, $p = .75$; see Figure 6).

Intergroup Behavior

Resource Allocation

Replicating prior work from Rhodes et al. (2017), children who heard generic language about Zarpies allocated resources (e.g., cookies) to Zarpies less often than children who heard specific language ($B = -.48$, $SE = .23$, $z = -2.09$, $p = .04$; see

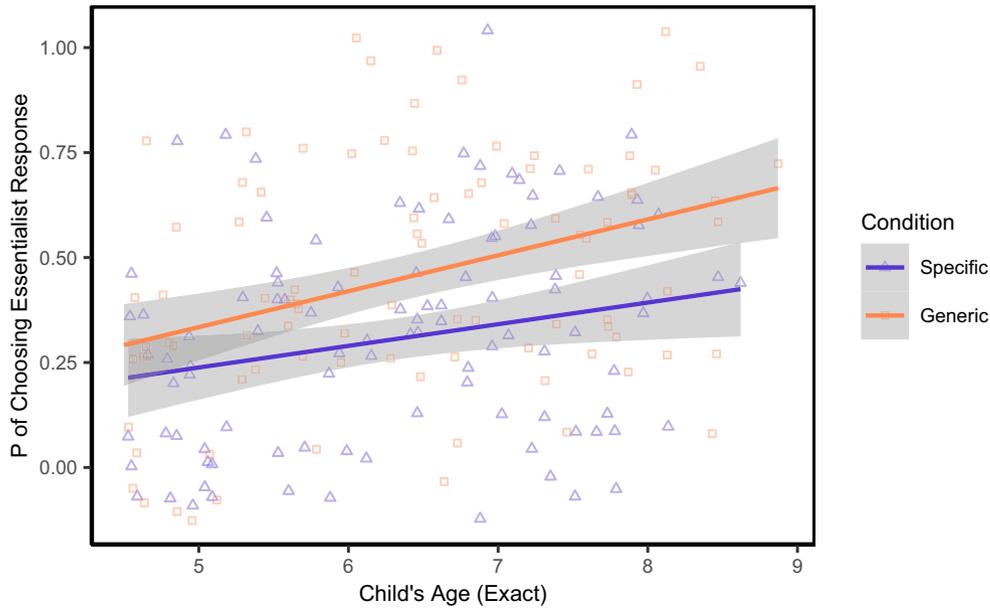


Figure 4. Participants' responses on the explanation items as a function of condition and age. Small shapes represent the proportion of essentialist responses given by each participant. Shaded regions represent 95% confidence intervals.

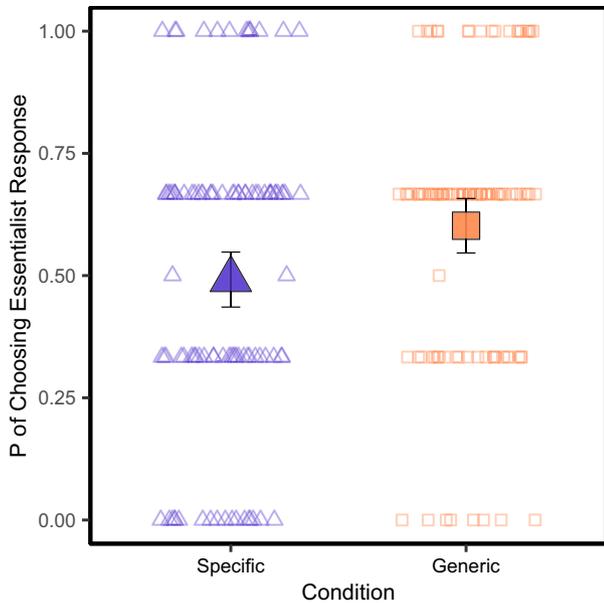


Figure 5. Participants' responses on the flexibility items as a function of condition ($M_{\text{specific}} = .49$; $M_{\text{generic}} = .60$). Smaller shapes represent the proportion of essentialist responses given by each participant. Larger shapes represent the mean proportion of essentialist responses by condition; error bars represent 95% confidence intervals around these means.

Figure 7). Allocation decisions did not vary by age ($B = -.12$, $SE = .14$, $z = -0.87$, $p = .39$), nor was there an interaction between age and condition ($B = .06$, $SE = .20$, $z = 0.28$, $p = .78$).

Expectations of Reciprocity

Our novel measure of expectations of reciprocity did not vary as a function of condition ($B = -.22$, $SE = .23$, $z = -0.97$, $p = .33$), age ($B = -.23$, $SE = .14$, $z = -1.63$, $p = .10$), or their interaction ($B = .36$, $SE = .20$, $z = 1.84$, $p = .07$).

Exploratory Analyses of Geographic Variables

Given the breadth of neighborhood-level diversity present in our sample (i.e., with respect to population size, racial-ethnic diversity, and household income) and the documented effects of contextual factors on essentialism (Mandalaywala et al., 2019; Rhodes & Gelman, 2009), we probed the influence of these neighborhood metrics in a set of exploratory (non-preregistered) analyses. Specifically, in a series of three exploratory analyses, we separately tested whether each of these metrics influenced composite levels of essentialism (or interacted with experimental condition). Across all three metrics of neighborhood-level diversity, the main effect of generic language on essentialism persisted ($ps = .003-.012$), providing evidence of the generalizability of our effect. Intriguingly, we found a separate main effect of neighborhood racial-ethnic diversity on children's level of essentialism, whereby greater racial-ethnic diversity in children's neighborhoods was associated with lower levels of

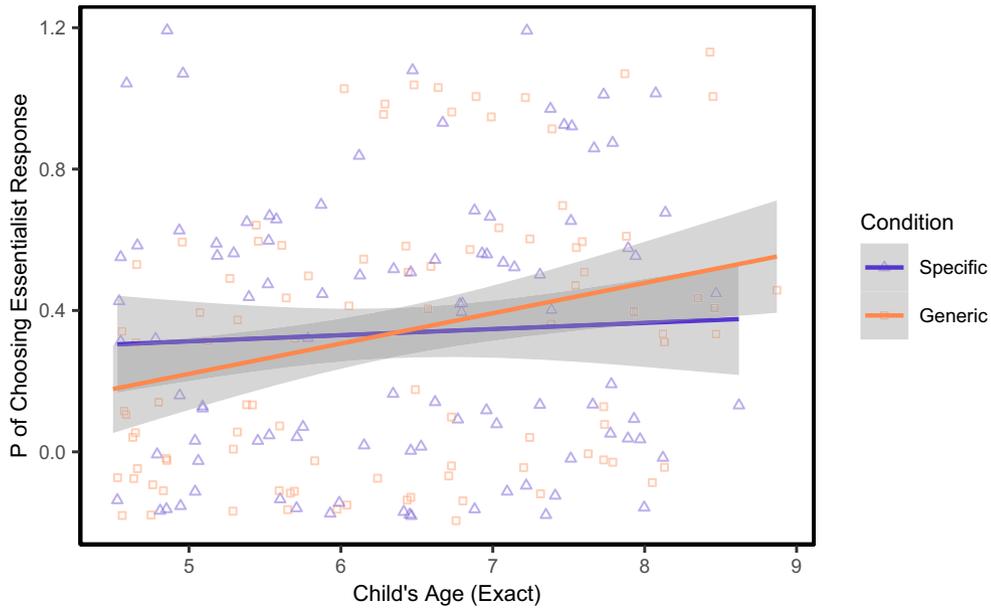


Figure 6. Participants' responses on the inheritance items as a function of condition and age. Small shapes represent the proportion of essentialist responses given by each participant. Shaded regions represent 95% confidence intervals.

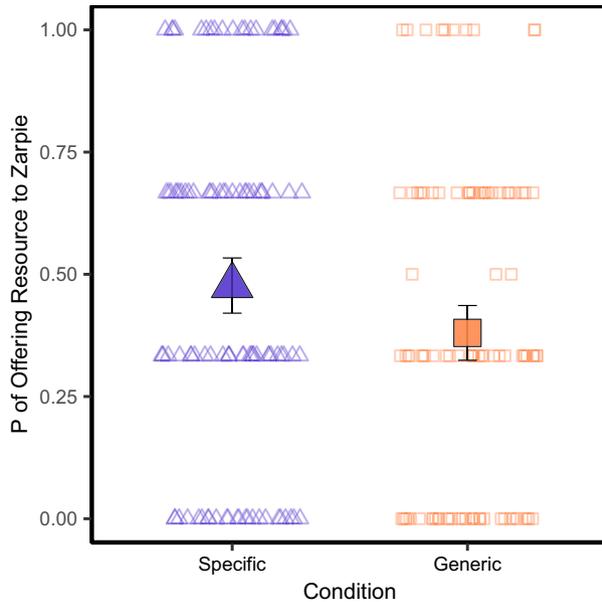


Figure 7. Participants' likelihood of allocating resources to Zarpies as a function of condition ($M_{\text{specific}} = .48$; $M_{\text{generic}} = .38$). Smaller shapes represent the likelihood of allocating resources to Zarpies for each participant. Larger shapes represent the mean likelihood of offering resources to Zarpies by condition; error bars represent 95% confidence intervals around these means.

essentialism overall, regardless of condition ($B = -.88$, $SE = .45$, $z = -1.94$, $p = .05$). Whereas previous work has found a specific relation

between exposure to racial diversity and racial essentialism in particular (Mandalaywala et al., 2019; Pauker, Tai, & Ansari, 2020; Rhodes & Gelman, 2009), these findings are the first that we are aware of to suggest possible effects of exposure to diversity on children's more general tendency to develop essentialist beliefs about a new social category. For more details on these exploratory analyses and consideration of other neighborhood variables see Supporting Information.

Discussion

This study found that commonplace patterns in language shape the development of children's essentialist beliefs. Children's inferences went far beyond the content of the language that they heard. For example, there is nothing in the sentence, "Zarpies hop over puddles" that directly communicates that a particular Zarpie does so for intrinsic, category-based reasons (e.g., because many Zarpies hate to get wet, instead of simply because he needed to get to the other side) or that this is the *only* way he would deal with a puddle (e.g., that he wouldn't sometimes choose to splash in it instead). Yet, generics—more so than descriptions of specific individuals—increased the likelihood that children would draw these types of inferences.

In this study, the implications of language for children's beliefs were similar but not identical across measure and age. Overall, generic language led children to endorse more intrinsic, category-related explanations and to view category boundaries as more rigid, but not to perceive the features described with generics as inherited (as assessed with the "switched-at-birth" paradigm; however, an exploratory analysis suggested effects on this measure might emerge with age). With respect to our null findings on this last measure, it is possible that children (and younger children in particular) had difficulty comprehending the switched-at-birth paradigm as administered here because, unlike previous studies (e.g., Hirschfeld, 1996; Rhodes et al., 2012; Taylor, 1996; Waxman et al., 2007), it did not include embedded comprehension questions and opportunities for the story to be repeated. It is also possible, however, that generic language on its own is not sufficient to lead children to think that category features are inherited, perhaps particularly when children are making judgments about new social groups, as done here (see Foster-Hanson et al., 2019; Gelman et al., 2010). From this perspective, generic language might shape beliefs about heritability only under certain circumstances; for instance, Noyes and Keil (2019, 2020) found that generic language led children and adults to believe that category features were inherited only when they described biological (not culturally determined) properties. One possibility, therefore, is that generics lead children to think that features have intrinsic causes (as shown in the present explanation task; and see Cimpian & Markman, 2009, 2011; Cimpian & Salomon, 2014), but not necessarily that they are *inherited* unless they are accompanied by further information about biology stating that the features specifically have biological origins (as shown in Noyes & Keil, 2020) or the preconceived notion of a category as biologically determined (e.g., as may be the case for gender categories, see Taylor et al., 2009).

This study found clear effects on two other components of essentialist beliefs—children's endorsements of intrinsic explanations and their beliefs that category-related properties are inflexible. Why do children draw these inferences from generics? These are not merely logical interpretations of the generic descriptions. For example, "Zarpies climb tall fences" is compatible with both the interpretation that Zarpies do so because they love to climb and the interpretation that they do so because their town is surrounded by a fence with no door, so they do not have another choice (Vasileyva et al.,

2018). Similarly, "Zarpies climb tall fences" could mean that they always and only do that, or that they sometimes climb fences and sometimes walk around them or open their doors.

We suggest that generics elicit essentialism—despite this ambiguity around their meaning—because of how they interact with the conceptual biases that underlie how people (and perhaps especially, young children) make sense of the world. From this perspective, young children recognize that there are many possible ways of classifying people, but view some of these classifications as holding more conceptual weight than others; that is, they expect some ways of categorizing people to mark coherent and objective clusters of people (e.g., gender groups) but other ways of classifying people to be more flexible and context-dependent (e.g., color groups in a classroom; for a review, see Rhodes & Mandalaywala, 2017). With this expectation in mind, children are on the lookout for cues in their environment as to which ways of classifying people are most stable and meaningful. Generic language provides a powerful cue. By two and a half years of age, children both recognize that generics refer to categories (rather than specific individuals; Gelman & Raman, 2003) and expect that adults know the *right* ways to label and categorize things (Jaswal & Neely, 2006). Thus, generic claims (e.g., "Zarpies hop over fences") can lead children to think that particular groupings (e.g., Zarpies) are informative for making sense of what different people are like because they signal that adults in their community see these categories as meaningful. In this way, generics do not *create* essentialist thought in children—they cue its application by serving as a signal of how their community sees the world.

Consistent with this perspective, children respond to generics differently when they come from knowledgeable versus unknowledgeable sources (Moty & Rhodes, 2019). Also, generics cue essentialist beliefs even when *all* they do is signal speakers' beliefs that the category is informative—even if they fail to communicate any of the features that category members have in common (Foster-Hanson et al., 2019). Of course, generics are not the only way to elicit essentialism; children might develop essentialist beliefs from their own observations or from other languages that directly communicate that properties of a category are intrinsic and shared across members. Nevertheless, generics appear to serve as particularly efficient cues to which types of categories are considered coherent and meaningful within a child's social environment.

This study also corroborated earlier findings of the consequences of subtle features of language for intergroup behavior. As in Rhodes et al. (2017), here generic language (compared to specific language) decreased children's willingness to allocate resources to a member of the Zarpie group. Why exactly generics have this effect on resource allocation remains unclear. Rhodes et al. (2017) proposed that perhaps essentialist beliefs (as triggered by generic language) increase the salience of the category boundary (i.e., between the child participant and Zarpies) as children consider future opportunities for social interaction and reciprocity; by this account, essentialism leads children to consider that members of another category might be unlikely to share with them, thus decreasing their own willingness to share. In this study, however, generics influenced only children's own willingness to share, not their expectations about who would share with them. Thus, the mechanisms by which generic language and essentialist beliefs shape children's intergroup behaviors will need to be examined in more detail in future work.

This study suggests that avoiding generic language may be one way to limit the spread of essentialist beliefs and avoid some of their negative consequences for intergroup relations (Levy & Dweck, 1999; Pauker, Ambady, & Apfelbaum, 2010). Yet, it is important not to misinterpret this work as advocating for a "color-blind" approach to social differences (Firebaugh & Davis, 1998; Plaut, 2002). Note that category labels were given in both conditions here—thus, this work does not suggest that adults should avoid labeling social differences or discussing aspects of social identity that are important in modern society and to people's lives. Instead, we argue that the form used in these types of communications is consequential and can lead children to draw inferences that adult communicators might not intend. Knowing that, in response to generic language, children tend to assume intrinsic causes (as shown by the present explanation items) and infer fixed relationships between categories and features (as shown by the flexibility items) means that these beliefs might need to be addressed directly in discussing social differences with children. For instance, adults may seek to make explicit the historical and structural forces that give rise to social differences, the process by which these forces were created in the first place, and the living reality that differences are not innate or inevitable (e.g., Noyes & Keil, 2019; Saul, 2017; Vasilyeva et al., 2018).

Unmoderated Remote Research as a Tool for Developmental Science

This project also illustrated how unmoderated remote research can help advance developmental science (Rhodes, Rizzo, et al., 2020). In this work, we used broadly available tools (that do not require programming expertise) to create simple narrated animations of the entire study, including the manipulation and test questions, that allowed children to complete the study independently. This approach facilitates reliability and replication. For example, all of our animated and narrated stimuli are all available on <https://osf.io/acrwq/>, and could be downloaded and used by other researchers in this area (even when testing in-person instead of online), thus making sure that methods are implemented in similar ways across studies and laboratories. Furthermore, this approach eliminates concerns about experimenter mistakes during testing, as well as about possible experimenter biases (in this study, parents—who sat next to their children during the study—were naturally blind to hypotheses, but if a researcher administered the study using these materials in the future, they could remain blind to which language condition the child received simply by having the child wear headphones during the manipulation phase). If this approach to implementing simple behavioral experiments with children were adopted broadly, it would increase the robustness of these experiments and facilitate the sharing of methods and stimuli for cross-lab replication—a practice that can be challenging for studies that involve elaborate scripted interactions with children that are difficult to fully describe in journal articles (for discussion, see Adolph, Gilmore, & Kennedy, 2017; Gilmore, Kennedy, & Adolph, 2018).

In addition to enhancing standardization and reproducibility, this approach offers the potential for developmental researchers to efficiently conduct studies that yield more generalizable findings by facilitating sampling with larger and more diverse populations (see Button et al., 2013; Frank, 2019; Frank et al., 2017; Nielson, Haun, Kartner, & Legare, 2017). Doing so allows testing new theoretical questions about how context, identity, and experience shape development—a particularly important issue when researchers are, as we are here, interested in children's understanding of social diversity itself—as well as more fine-grained analyses of how effects change across age, allowing for new insights into developmental processes. Yet, to capitalize

more fully on this potential, future research will be needed to explore how to further increase the accessibility of remote online studies for families from more racially and ethnically diverse backgrounds, as well as for families from lower-income communities and with more limited access to technology (Lourenco & Tasimi, 2020).

Although remote research is certainly not possible for all studies (e.g., those of motor behavior, Adolph, Vereijken, & Shrout, 2003, or that require social interaction with an experimenter, Foster-Hanson et al., 2019), the present method builds on several others (Scott & Schulz, 2017; Scott et al., 2017; Sheskin et al., 2020) showing the feasibility of conducting developmental research online. We add to this small (but proliferating) body of studies by highlighting the importance of this approach for research on social cognition in particular, by showing how this approach can be implemented using tools that are broadly available (without extensive programming knowledge) and in a manner that children can complete on their own (without directly interacting with an experimenter via video-conference, a common protocol for online research; see Sheskin et al., 2020), and by extending this approach to more complicated and longer experimental paradigms than have been previously tried with children online. Given recent efforts at replication in related fields (Camerer et al., 2018), the field needs new approaches for conducting experimental developmental science with larger samples and using methods that are easier to share and replicate (for recent efforts in this area, see Chouinard, Scott, & Cusack, 2019; Databrary, 2012; Frank et al., 2017; Rubio-Fernández, 2019; Scott & Schulz, 2017; Wood & Wood, 2019). We think that online research, paired with other best practices for promoting robust science, including preregistration, ensuring sufficient power, and methods and data sharing (including videos of experimental sessions via Databrary, as we have done here; Rhodes, 2020), has great potential for moving developmental science forward.

References

- Adolph, K. E., Gilmore, R. O., & Kennedy, J. L. (2017). *Video data and documentation will improve psychological science*. Psychological Science Agenda. Retrieved from <https://www.apa.org/science/about/psa/2017/10/index.aspx>
- Adolph, K. E., Vereijken, B., & Shrout, P. E. (2003). What changes in infant walking and why. *Child Development*, 74, 475–497. <https://doi.org/10.1111/1467-8624.7402011>
- Allport, G. W. (1954). *The nature of prejudice*. Cambridge, MA: Perseus.
- Astuti, R., Solomon, G. E. A., & Carey, S. (2004). Constraints on conceptual development. *Monographs of the Society for Research in Child Development*, 69, 1–135. <https://doi.org/10.1111/j.0037-976X.2004.00296.x>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Birnbaum, D., Deeb, I., Segall, G., Ben-Eliyahu, A., & Diesendruck, G. (2010). The development of social essentialism: The case of Israeli children's inferences about Jews and Arabs. *Child Development*, 81, 757–777. <https://doi.org/10.1111/j.1467-8624.2010.01432.x>
- Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, 14, 365–376. <https://doi.org/10.1038/nrn3475>
- Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., . . . Wu, H. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour*. <https://doi.org/10.1038/s41562-018-0399-z>
- Chouinard, B., Scott, K., & Cusack, R. (2019). Using automatic face analysis to score infant behaviour from video collected online. *Infant Behavior and Development*. <https://doi.org/10.1016/j.infbeh.2018.11.004>
- 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/non-generic distinction influences how children interpret new information about social others. *Child Development*, 82, 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, 461–480. <https://doi.org/10.1017/S0140525X13002197>
- Databrary. (2012). *The Databrary project: A video data library for developmental science*. New York, NY: New York University. Retrieved from <http://databrary.org>
- Deeb, I., Segall, G., Birnbaum, D., Ben-Eliyahu, A., & Diesendruck, G. (2011). Seeing isn't believing: The effect of intergroup exposure on children's essentialist beliefs about ethnic categories. *Journal of Personality & Social Psychology*, 101, 1139–1156. <https://doi.org/10.1037/a0026107>
- Diesendruck, G., Goldfein-Elbaz, R., Rhodes, M., Gelman, S., & Neumark, N. (2015). "Cross-cultural differences in

- children's beliefs about the objectivity of social categories": Erratum. *Child Development*, 86, 659. <https://doi.org/10.1111/cdev.12369>
- Diesendruck, G., & Menahem, R. (2015). Essentialism promotes children's inter-ethnic bias. *Frontiers in Psychology*, 6, Article 1180. <https://doi.org/10.3389/fpsyg.2015.01180>
- Firebaugh, G., & Davis, K. E. (1998). Trends in antiblack prejudice, 1972–1984: Region and cohort effects. *American Journal of Sociology*, 94, 251–272. <https://doi.org/10.1086/228991>
- Foster-Hanson, E., Leslie, S. J., & Rhodes, M. (2019, January 20). *Speaking of kinds: How generic language shapes the development of category representations*. <https://doi.org/10.31234/osf.io/28qf7>
- Foster-Hanson, E., & Rhodes, M. (2019). Is the most representative skunk the average or the stinkiest? Development changes in representation of biological categories. *Cognitive Psychology*, 110, 1–15.
- Frank, M. C. (2019). Towards a more robust and replicable science of infant development. *Infant Behavior and Development*, 57, 101349. <https://doi.org/10.1016/j.infbeh.2019.101349>
- Frank, M. C., Bergelson, E., Bergmann, C., Cristia, A., Floccia, C., Gervain, J., . . . Yurovsky, D. (2017). A collaborative approach to infant research: Promoting reproducibility, best practices, and theory-building. *Infancy*. <https://doi.org/10.1111/inf.12182>
- Gaither, S. E., Schultz, J. R., Pauker, K., Sommers, S. R., Maddox, K. B., & Ambady, N. (2014). Essentialist thinking predicts decrements in children's memory for racially ambiguous faces. *Developmental Psychology*, 50, 482. <https://doi.org/10.1037/a0033493>
- Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. New York, NY: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195154061.001.0001>
- Gelman, S. A., Coley, J. D., Rosengren, K. S., Hartman, E., Pappas, A., & Keil, F. C. (1998). Beyond labeling: The role of maternal input in the acquisition of richly-structured categories. *Monographs of the Society for Research in Child Development*, 63(1, Serial No. 253). <https://doi.org/10.2307/1166211>
- Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition*, 23, 183–209. [https://doi.org/10.1016/0010-0277\(86\)90034-X](https://doi.org/10.1016/0010-0277(86)90034-X)
- Gelman, S. A., & Raman, L. (2003). Preschool children use linguistic form class and pragmatic cues to interpret generics. *Child Development*, 74, 308–325. <https://doi.org/10.1111/1467-8624.00537>
- Gelman, S. A., & Rhodes, M. (2012). "Two-thousand years of stasis": How psychological essentialism impedes evolutionary understanding. In *Evolution challenges: Integrating research and practice in teaching and learning about evolution*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199730421.003.0001>
- Gelman, S. A., & Roberts, S. O. (2017). How language shapes the cultural inheritance of categories. *Proceedings of the National Academy of Sciences of the United States of America*, 114, 7900–7907. <https://doi.org/10.1073/pnas.1621073114>
- Gelman, S. A., & Tardif, T. (1998). A cross-linguistic comparison of generic noun phrases in English and Mandarin. *Cognition*, 66, 215–248. [https://doi.org/10.1016/S0010-0277\(98\)00021-3](https://doi.org/10.1016/S0010-0277(98)00021-3)
- Gelman, S. A., Taylor, M. G., & Nguyen, S. P. (2004). Mother-child conversations about gender: Understanding the acquisition of essentialist beliefs. *Monographs of the Society for Research in Child Development*, 69(Serial No. 275), i–142.
- Gelman, S. A., Ware, E. A., & Kleinberg, F. (2010). Effects of generic language on category content and structure. *Cognitive Psychology*, 61, 273–301. <https://doi.org/10.1016/j.cogpsych.2010.06.001>
- Gelman, S. A., & Wellman, H. M. (1991). Insides and essences: Early understandings of the non-obvious. *Cognition*. [https://doi.org/10.1016/0010-0277\(91\)90007-Q](https://doi.org/10.1016/0010-0277(91)90007-Q)
- Gilmore, R. O., Kennedy, J. L., & Adolph, K. E. (2018). Practical solutions for sharing data and materials from psychological research. *Advances in Methods and Practices in Psychological Science*. <https://doi.org/10.1177/2515245917746500>
- Haslam, N., Rothschild, L., & Ernst, D. (2000). Essentialist beliefs about social categories. *British Journal of Social Psychology*, 39, 113–127. <https://doi.org/10.1348/014466600164363>
- Haslam, N., Rothschild, L., & Ernst, D. (2002). Are essentialist beliefs associated with prejudice? *British Journal of Social Psychology*, 41, 87–100. <https://doi.org/10.1348/014466602165072>
- Hirschfeld, L. A. (1996). *Race in the making: Cognition, culture, and the child's construction of human kinds*. Cambridge, MA: MIT Press.
- Jaswal, V. K., & Neely, L. A. (2006). Adults don't always know best: Preschoolers use past reliability over age when learning new words. *Psychological Science*, 17, 757–758. <https://doi.org/10.1111/j.1467-9280.2006.01778.x>
- Kinzler, K. D., & Dautel, J. B. (2012). Children's essentialist reasoning about language and race. *Developmental Science*, 15, 131–138. <https://doi.org/10.1111/j.1467-7687.2011.01101.x>
- Leslie, S. J. (2013). Essence and natural kinds: When science meets preschooler intuition. *Oxford Studies in Epistemology*, 4, 108–165.
- Leslie, S. J. (2017). The original sin of cognition: Fear, prejudice and generalization. *The Journal of Philosophy*, 114, 393–421. <https://doi.org/10.5840/jphil2017114828>
- Levy, S. R., & Dweck, C. S. (1999). The impact of children's static versus dynamic conceptions of people on stereotype formation. *Child Development*, 70, 1163–1180. <https://doi.org/10.1111/1467-8624.00085>
- Lingeman, J., Freeman, C., & Adolph, K. E. (2014). *Datavyu* (Version 1.2.2) [Software]. Retrieved from <http://datavyu.org>

- Lourenco, S. F., & Tasimi, A. (2020, May 7). *No participant left behind: Conducting science during COVID-19*. <https://doi.org/10.31234/osf.io/r62gc>
- Mandalaywala, T. M., Amodio, D., & Rhodes, M. (2017). Essentialism promotes racial prejudice by increasing endorsement of social hierarchies. *Social Psychological and Personality Science*, *9*, 461–469. <https://doi.org/10.1177/1948550617707020>
- Mandalaywala, T. M., Fleener, C., & Maestriperieri, D. (2015). Intelligence in nonhuman primates. In S. Goldstein, D. Princiotto, & J. A. Naglieri (Eds.), *Handbook of intelligence* (pp. 27–46). New York, NY: Springer.
- 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*, 437–453. <https://doi.org/10.1111/cdev.13008>
- Mayr, E. (1982). *The growth of biological thought: Diversity, evolution, and inheritance*. Cambridge, MA: Belknap Press.
- Mayr, E. (1988). *Toward a new philosophy of biology, observations of an evolutionist*. Cambridge, MA: Harvard University Press.
- Moty, K., & Rhodes, M. (2019, August 29). *The unintended consequences of the things we say: What generics communicate to children about unmentioned categories*. <https://doi.org/10.31234/osf.io/zkjjr>
- Nielsen, M., Haun, D., Kärtner, J., & Legare, C. H. (2017). The persistent sampling bias in developmental psychology: A call to action. *Journal of Experimental Child Psychology*, *162*, 31–38. <https://doi.org/10.1016/j.jecp.2017.04.017>
- Noyes, A., & Keil, F. C. (2019). Generics designate kinds but not always essences. *Proceedings of the National Academy of Sciences of the United States of America*, *116*, 20354–20359. <https://doi.org/10.1073/pnas.1900105116>
- Noyes, A., & Keil, F. C. (2020). There is no privileged link between kinds and essences early in development. *Proceedings of the National Academy of Sciences of the United States of America*, *117*, 10633–10635. <https://doi.org/10.1073/pnas.2003627117>
- Pauker, K., Ambady, N., & Apfelbaum, E. (2010). Race salience and essentialist thinking in racial stereotype development. *Child Development*, *81*, 1799–1813. <https://doi.org/10.1111/j.1467-8624.2010.01511.x>
- Pauker, K., Tai, C., & Ansari, S. (2020). *Contextualizing the development of social essentialism*. <https://doi.org/10.31234/osf.io/6e3th>
- Pehrson, S., Brown, R., & Zagefka, H. (2009). When does national identification lead to the rejection of immigrants? Cross-sectional and longitudinal evidence for the role of essentialist in-group definitions. *British Journal of Social Psychology*, *48*(1), 61–76. <https://doi.org/10.1348/014466608X288827>
- Plaut, V. C. (2002). Cultural models of diversity in America: The psychology of difference and inclusion. In R. Shweder, M. Minow, & H. R. Markus (Eds.), *Engaging cultural differences: The multicultural challenge in aliberal democracy* (pp. 365–395). New York, NY: Russell Sage.
- Prentice, D. A., & Miller, D. T. (2007). Psychological essentialism of human categories. *Current Directions in Psychological Science*, *16*, 202–206. <https://doi.org/10.1111/j.1467-8721.2007.00504.x>
- Rhodes, M. (2020). *Does it matter how we speak about social kinds? A large, pre-registered, online experimental study of how language shapes the development of essentialist beliefs*. Databrary. Retrieved from <https://nyu.databrary.org/volume/1159>
- 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 of the United States of America*, *117*, 9808–9814. <https://doi.org/10.1073/pnas.1919646117>
- Rhodes, M., & Gelman, S. A. (2009). A developmental examination of the conceptual structure of animal, artifact, and human social categories across two cultural contexts. *Cognitive Psychology*, *59*, 244–274. <https://doi.org/10.1016/j.cogpsych.2009.05.001>
- Rhodes, M., Gelman, S. A., & Karuza, J. C. (2014). Pre-school ontology: The role of beliefs about category boundaries in early categorization. *Journal of Cognition and Development*, *15*, 78–93. <https://doi.org/10.1080/15248372.2012.713875>
- Rhodes, M., Leslie, S. J., Saunders, K., Dunham, Y., & Cimpian, A. (2017). How does social essentialism affect the development of inter-group relations? *Developmental Science*, *21*, e12509. <https://doi.org/10.1080/15248372.2012.713875>
- Rhodes, M., Leslie, S. J., & Tworek, C. M. (2012). Cultural transmission of social essentialism. *Proceedings of the National Academy of Sciences*, *109*, 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*, e1437. <https://doi.org/10.1002/wcs.1437>
- Rhodes, M., & Moty, K. (2020). What is social essentialism and how does it develop? In *Advances in child development and behavior* (Vol. 59, pp. 1–30). Elsevier. <https://doi.org/10.1016/bs.acdb.2020.05.001>
- Rhodes, M., Rizzo, M. T., Foster-Hanson, E., Moty, K., Leshin, R. A., Wang, M., . . . Ocampo, J. D. (2020). Advancing developmental science via unmoderated remote research with children. *Journal of Cognition and Development*, *21*, 477–493. <https://doi.org/10.1080/15248372.2020.1797751>
- Rizzo, M. T., Elenbaas, L., & Vanderbilt, K. E. (2020). Do children distinguish between resource inequalities with individual versus structural origins? *Child Development*, *91*, 439–455. <https://doi.org/10.1111/cdev.13181>
- Roberts, S. O., & Gelman, S. A. (2016). Can White children grow up to be Black? Children’s reasoning about the stability of emotion and race. *Developmental Psychology*, *52*, 887–893. <https://doi.org/10.1037/dev0000132>
- Roberts, S., Ho, A., Rhodes, M., & Gelman, S. A. (2017). Making boundaries great again? Essentialism and

- support for boundary enhancing initiatives. *Personality and Social Psychology Bulletin*, 43, 1643–1658. <https://doi.org/10.1177/0146167217724801>
- Rothbart, M., & Taylor, M. (1992). Category labels and social reality: Do we view social categories as natural kinds? In G. R. Semin & K. Fiedler (Eds.), *Language, interaction and social cognition* (pp. 11–36). Thousand Oaks, CA: Sage Publications, Inc.
- Rubio-Fernández, P. (2019). Publication standards in infancy research: Three ways to make Violation-of-Expectation studies more reliable. *Infant Behavior and Development*. <https://doi.org/10.1016/j.infbeh.2018.09.009>
- Saul, J. (2017). Implicit bias, stereotype threat, and epistemic injustice. In *The Routledge handbook of epistemic injustice*. <https://doi.org/10.4324/9781315212043>
- Scott, K., Chu, J., & Schulz, L. (2017). Lookit (Part 2): Assessing the viability of online development research, results from three case studies. *Open Mind: Discoveries in Cognitive Science*, 1(1), 15–29. https://doi.org/10.1162/opmi_a_00001
- Scott, K., & Schulz, L. (2017). Lookit (Part 1): A new online platform for developmental research. *Open Mind: Discoveries in Cognitive Science*, 1(1), 4–14. https://doi.org/10.1162/opmi_a_00002
- Segall, G., Birnbaum, D., Deeb, I., & Diesendruck, G. (2015). The intergenerational transmission of ethnic essentialism: how parents talk counts the most. *Developmental Science*, 18, 543–555.
- Sheskin, M., Scott, K., Mills, C. M., Bergelson, E., Bonawitz, E., Spelke, E. S., Jara-Ettinger, J. (2020). Online developmental science to foster innovation, access, and impact. *Trends in Cognitive Sciences*, 24, 675–678. <https://doi.org/10.1016/j.tics.2020.06.004>
- Shtulman, A., & Schulz, L. (2008). The relation between essentialist beliefs and evolutionary reasoning. *Cognitive Science*, 32, 1049–1062. <https://doi.org/10.1080/0364021080189786>
- Smyth, K., Feeney, A., Eidson, R. C., & Coley, J. D. (2017). Development of essentialist thinking about religion categories in Northern Ireland (and the United States). *Developmental Psychology*. <https://doi.org/10.1037/dev0000253>
- 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/10.2307/1131718>
- 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, 461–481. <https://doi.org/10.1111/j.1467-8624.2009.01272.x>
- Vasilyeva, N., Gopnik, A., & Lombrozo, T. (2018). The development of structural thinking about social categories. *Developmental Psychology*. <https://doi.org/10.1037/dev0000555>
- 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*. <https://doi.org/10.1037/0012-1649.43.2.294>
- Wood, S. M. W., & Wood, J. N. (2019). Using automation to combat the replication crisis: A case study from controlled-rearing studies of newborn chicks. *Infant Behavior and Development*. <https://doi.org/10.1016/j.infbeh.2019.101329>
- Yzerbyt, V., Corneille, O., & Estrada, C. (2001). The interplay of subjective essentialism and entitativity in the formation of stereotypes. *Personality and Social Psychology Review*, 5, 141–155. https://doi.org/10.1207/S15327957PSPR0502_5

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's website:

Appendix S1. See Supplementary Online Materials for (I) analyses separated by country (United States vs. United Kingdom), (II) analyses excluding all trials identified as involving parental interference, (III) descriptive analyses of parent-child interactions during the study session, (IV) descriptive analyses of neighborhood-level variables, and (V) inferential analyses of neighborhood-level variables