So it is, so it shall be: Group regularities license children's prescriptive judgments

Steven O. Roberts, **Susan A. Gelman**, and **Arnold K. Ho**

Department of Psychology, University of Michigan

Abstract

When do descriptive regularities (what characteristics individuals have) become prescriptive norms (what characteristics individuals *should* have)? We examined children's (4–13 years) and adults' use of group regularities to make prescriptive judgments, employing novel groups (Hibbles and Glerks) that engaged in morally neutral behaviors (e.g., eating different kinds of berries). Participants were introduced to conforming or non-conforming individuals (e.g., a Hibble who ate berries more typical of a Glerk). Children negatively evaluated non-conformity, with negative evaluations declining with age (Study 1). These effects were replicable across competitive and cooperative intergroup contexts (Study 2), and stemmed from reasoning about group regularities rather than reasoning about individual regularities (Study 3). These data provide new insights into children's group concepts and have important implications for understanding the development of stereotyping and norm enforcement.

Keywords

cognitive development; social cognition; novel groups; group norms; norm enforcement; conformity

Social categories entail descriptive regularities (i.e., characteristics shared by individuals within the group): Doctors are expected to help others, boys are expected not to wear lipstick, and in some contexts, Black students are expected to underachieve academically (Devine, 1989; Kalish, 2012). Social categories may also entail prescriptive norms (i.e., characteristics that individuals within the group *should* have): Doctors who don't help others, boys who wear lipstick, or Black students who excel academically may be evaluated negatively (Blakemore, 2003; Durkee & Williams, 2013; Levy, Taylor, & Gelman, 1995). In the present research, we examined the extent to which descriptive regularities license prescriptive judgments. This research has important implications for a variety of disciplines, ranging from comparative biology to philosophy, by providing insight into the cognitive mechanisms that contribute to a cross-species preference for conformity (Claidière &

Whiten, 2012). We take a developmental approach, as this method permits identifying early-emerging, foundational processes that underlie the more complex representations held by adults (Gelman, 2003; Olson & Dweck, 2008; Rhodes, 2012; Wellman, 2014).

Adults and young children alike make use of group concepts to infer regularities regarding familiar categories such as gender and race, as well as novel groups to which they themselves do not belong (Fiske, Cuddy, Glick, & Xu, 2002; Gelman, Ware, & Kleinberg, 2010; Liben, Bigler, & Krogh, 2002; Roberts & Gelman, 2015; Shutts, Pemberton Roben, & Spelke, 2013). They also often treat regularities as prescriptive, at least under certain conditions. By 4 years of age, children disapprove of individuals who violate moral precepts or gender norms, and they treat individuals who follow descriptive regularities as more representative group members than those who do not (Blakemore, 2003; Cooley & Killen, 2015; Kalish, 2012; Levy et al., 1995; Rhodes & Chalik, 2013). By adulthood, the tendency to go from the descriptive to the prescriptive can have severe social consequences. For instance, the belief that women are collaborative and deferential, whereas men are independent and assertive, can lead to negativity toward individuals who violate those beliefs (e.g., a woman who prefers independence over collaboration), which ultimately, can produce workplace bias and prevent career advancement (for a review, see Heilman, 2012).

These studies yield important insights into prescriptive reasoning, yet in each case participants had access to cues beyond group membership per se that may have encouraged a prescriptive stance. Some studies focused on familiar groups and properties for which there is cultural input regarding non-conformity (e.g., Blakemore, 2003; Levy et al., 1995). For instance, from an early age, young children are taught that gender reflects real and meaningful differences (e.g., via gender segregated spaces and occupations), and they may even encounter punishment or ridicule from peers or adults when they attempt to transgress gender boundaries (e.g., Liben et al., 2001). Other studies involved participants' own group membership and may have therefore been influenced by in-group positivity (e.g., Mulvey, Hitti, Rutland, Abrams, & Killen, 2014). For instance, by 6 years of age, children perceive non-conforming in-group members more negatively than non-conforming out-group members, because the former threaten in-group cohesion and loyalty (Rutland, Hitti, Mulvey, Abrams, & Killen, 2015). Others studies included morally laden behaviors (e.g., harming others) that have inherent meaning to children as well as adults (e.g., Mulvey, in press; Smetana et al., 2012), or they primed inter-group competition which may have heightened attention toward groups (e.g., Rhodes, 2012). Thus, an important open question is the extent to which prescriptive judgments persist even without these extra cues--for example, when people are asked to reason about unfamiliar groups to which they do not belong, and regarding morally neutral behaviors in noncompetitive contexts.

Research on "over-imitation" suggests an early-emerging tendency to treat descriptive rules as prescriptive norms, even under the minimal conditions outlined above (see Nielsen & Haun, 2016). After 2-year-old children observed a behavior paired with a goal (e.g., twiddling an object before using it to open a box), they followed the behavior rigidly and expected others to do the same, even when they understood that the behavior itself was irrelevant for achieving the goal. When children saw someone omit the irrelevant behavior before achieving the goal, they protested using prescriptive language (e.g., saying that the

individual *should* follow the behavior and that it is *wrong* to do otherwise), and similarly, children who observed someone violate the rules of a newly learned game responded with a prescriptive critique (Kenward, 2012; Rakoczy & Schmidt, 2013). Taken together, these findings suggest that children take a prescriptive interpretation of regularities that guides what they believe individuals should do.

Of course, prescriptive judgments regarding imitation or games need not extend to prescriptive judgments toward social groups. Yet children may similarly treat group regularities as prescriptive and thus negatively evaluate non-conforming individuals. An evolutionary, functional account of "norm psychology"—arguing that adherence to norms and punishment of non-conformity facilitates cultural learning and group coordination— supports this possibility (e.g., Chudek & Henrich, 2011), as does the tendency for humans (and non-human animals) to engage in group conformity (Cialdini & Goldstein, 2004; Claidière & Whiten, 2012; Haun & Tomasello, 2011; Schillaci & Kelemen, 2014). Alternatively, individuals may initially treat social groups as entailing strictly descriptive regularities, and in the absence of experiences with the group, membership in the group, or morally laden behaviors, they may not interpret those regularities as prescriptive.

We examined how people used information about the way a group *is* (i.e., descriptive regularities) to make inferences about the way individuals *should* be (i.e., prescriptive judgments), under conditions with novel groups, morally neutral and harmless behaviors, and an absence of in-group membership or competition cues. Exploring this question is important, as a link between descriptive regularities and prescriptive judgments under such "minimal conditions" may reveal how readily adapted humans are to engage in norm enforcement and conformity-based reasoning. That is, once we acquire the belief that group concepts are diagnostic of individual characteristics (i.e., stereotypes) we may have a cognitive basis from which to negatively evaluate non-conforming individuals.

The Present Research

In three experiments, we introduced participants to two novel groups, Hibbles and Glerks, who were described in terms of morally neutral regularities (e.g., the kinds of berries they eat). Participants were then introduced to conforming or non-conforming individuals. We measured participants' *disapproval* or *approval* of the behavior, *degree of negativity* toward behaviors they disapproved of, and open-ended *explanations* regarding how they justified their evaluations. This research differs from previous research in three important ways. First, we did not ask about familiar groups or behaviors about which participants have prior expectations (e.g., wearing gendered clothing). Second, we did not ask about behaviors with moral underpinnings (e.g., harming others). Third, participants were not members of either of the groups.

STUDY 1

Study 1 focused on children's and adults' evaluations of conforming and non-conforming individuals. We focused on children from 4–13 years of age, because important changes are taking place over this period in children's experiences in school settings and peer group

interactions, as well as in their social categorization, identity, essentialism, and stereotypes (e.g., they become more sophisticated in social categorization and attuned to the importance of social categories) and because this is a common age range in this kind of research; Bigler and Liben, 1997; Quintana, 1998; Rhodes & Gelman, 2009; Roberts & Gelman, 2015, 2016a, 2016b). Adults provide an important comparison group to assess mature patterns of reasoning.

Method

Participants—Four age groups of participants were included ($N = 106$): twenty-four 4- to 6-year-olds (39% female, M age = 5.37 years, $SD = .80$), thirty-one 7- to 9-year-olds (52% female, M age = 8.52 years, $SD = .90$), twenty-seven 10- to 13-year-olds (33% female, M age = 11.39 years, $SD = .97$), and twenty-four adults (46% female, M age = 20.87, $SD = 1.65$). Children were recruited in the Midwest at two university-affiliated museums. Adults were recruited on a college campus. The sample was mostly white/European American (67% white/European American, 13% Asian American/Asian/Pacific Islander, 6% Hispanic/ Latino, 6% Multiracial, 4% black/African American, and 4% other or not reported).[1]

Materials and procedure—All materials were presented on a laptop using PowerPoint. First, participants were introduced to cartoon drawings depicting members of two novel groups: Hibbles and Glerks, and were told, "I'm going to tell you about two kinds of groups. This group (*pointing*) is called Hibbles, and this group (*pointing*) is called Glerks." Each group consisted of three individuals located on one side of the screen (left or right), and group membership was portrayed by clothing pattern (green stripes, orange triangles) and group labels (Hibbles, Glerks). Following the introduction of the novel groups, participants received eight test trials; for each, they heard a property of each group and were then introduced to a conforming or non-conforming individual. Across the eight trials, there were four behavioral domains: Food ("Hibbles eat these kinds of berries [*pointing*] and Glerks eat these kinds of berries [*pointing*]. Look [*revealing the target*], this Glerk is eating these kinds of berries [*pointing*]"), Games ("Hibbles play games with this kind of toy [*pointing*] and Glerks play games with this kind of toy [*pointing*]. Look [*revealing the target*], this Glerk is playing games with this kind of toy [*pointing*]"), Language ("Hibbles speak this kind of language [*pointing*] and Glerks speak this kind of language [*pointing*]. Look [*revealing the target*], this Glerk is speaking this kind of language [*pointing*]"), and Music ("Hibbles listen to this kind of music [*pointing*] and Glerks listen to this kind of music [*pointing*]. Look [*revealing the target*], this Glerk is listening to this kind of music").

Each behavior was depicted by a character that matched its respective group in color (e.g., orange clothing pattern corresponded with orange berries, an orange boomerang-shaped toy, orange words in a foreign script within a speech bubble, and an orange musical note). Across participants, we counterbalanced which pattern depicted which group, which pattern was associated with which group label, and the left-right order in which the groups were presented. The eight test trials were presented in random order. Half of the trials depicted a

---

[1]As with prior developmental research with comparable designs (e.g., Rhodes & Chalik, 2013; Roberts & Gelman, 2015), we sought to have at least 24 participants within each age group and condition (across all studies). Data collection was stopped the day this goal was achieved.

non-conforming individual and the other half depicted a conforming individual (for an example trial, see Figure S1 in the online supplemental materials).

Measures and coding—Across all studies, all independent variables and manipulations are reported. The first measure was the frequency with which participants evaluated behaviors as "okay" or "not okay" (evaluation, e.g., "Is it okay or not okay for this Glerk to eat these kinds of berries?"). We calculated the frequency with which these evaluations occurred for both conformity and non-conformity trials (potential range for each was 0 to 4). Because the frequencies of "okay" and "not okay" evaluations were precise inverses, we focused on "not okay" evaluations as the first dependent variable, which reflected disapproval toward a given behavior. The second measure was the negativity with which participants rated "not okay" behaviors. That is, participants who evaluated a behavior as "not okay" were presented a scale with three increasingly unhappy faces and asked, "Is it a little bad, pretty bad, or very, very bad?" (1 = *a little bad*, 2 = *pretty bad*, 3 = *very, very bad*).

The third measure was derived from the explanations that participants provided for their evaluation (i.e., approval or disapproval), which were recorded verbatim. Given the open-ended nature of these responses, participants generated a wide range of explanations. Using previous research as a theoretical guide (Rhodes, 2014; Rutland et al., 2015), explanations were coded into five types: (a) norm-based (e.g., "They are supposed to"; "They are not allowed to"), (b) group-based (e.g., "That's what Glerks do"; "They are the same kind"), (c) individual-based (e.g., "He wants to"; "Different people like different things"), (d) similarity-based (e.g., "They are orange"; "They look different"), and (e) other (e.g., "It has a weird face"). Participants could appeal to multiple explanations within a single response, as codes were not mutually exclusive. Two independent coders who were blind to the hypotheses of the study conducted the coding (Cohen's kappa = .76) with disagreements resolved by discussion. We then calculated the percentages of times that each type of explanation was provided (out of the total number of trials) for each response type. Importantly, each type of explanation's valence could have been positive or negative. For instance, for group-based or similarity-based explanations, they could have referred to shared group membership (e.g., "because they are a Hibble") or different group membership (e.g., "because they are not a Glerk"). Nevertheless, we used these broad codes because in each case participants referred to the same mechanism (e.g., the group, individuality, norms, similarity). For a similar coding scheme, see Rhodes (2014).

Results

There were no effects of the counterbalancing factors across any of the studies. Also, preliminary analyses revealed that all effects were robust across behavioral domains (i.e., food, games, language, music; for analyses by domain see Appendix A in the online supplemental materials). Accordingly, data were collapsed across these variables. All reported effects were followed by Bonferroni-corrected pairwise comparisons.

Disapproval toward conformity and non-conformity—We conducted a 4 (age group: 4–6, 7–9, 10–13, adult) × 2 (behavior: conformity, non-conformity) repeated measures ANOVA with age group as a between-subjects variable, behavior as the within-

subjects variable, and the frequency of not-okay evaluations as the dependent variable. Higher scores indicated a greater frequency of evaluating behaviors as "not okay", and therefore, greater disapproval (scores could range from 0 to 4). There were significant main effects of age group, $F(3, 102) = 22.12$, $p < .001$, $\eta_p^2 = .40$, and behavior, $F(1, 102) = 82.24$, $p < .001$, $\eta_p^2 = .45$, and a significant interaction of age group and behavior, $F(3, 102) = 6.10$, $p = .001$, $\eta_p^2 = .15$. All child groups, but not adults ($p = .44$), were significantly more disapproving of non-conformity than of conformity ($ps < .001$), and all child groups were more disapproving of non-conformity than adults ($ps < .001$). Also, 4- to 6-year-olds were more disapproving of conformity than all other age groups (who did not differ significantly from one another). These data are presented graphically in Figure 1.

We next conducted a series of one-sample t-tests to compare disapproval rates against chance (i.e., 2). For conformity, disapproval was below chance in all age groups: 4 to 6: $M = 1.13$, $SE = .16$, $t(23) = -2.73$, $p = .012$, $d = .56$, 7 to 9: $M = .07$, $SE = .14$, $t(30) = 30.00$, $p < .001$, $d = 5.38$; 10 to 13: $M = .04$, $SE = .15$, $t(26) = -53$, $p < .001$, $d = 10.19$; Adults: $M = 0$. For non-conformity, 4- to 6-year-olds were above chance, 7- to 9-year-olds and 10- to 13-year-olds were at chance, and adults were below chance: 4 to 6: $M = 3.04$, $SE = .32$, $t(23) = 3.43$, $p = .002$, $d = .72$, 7 to 9: $M = 2.29$, $SE = .28$, $t(30) = .867$, $p = .393$, $d = .16$, 10 to 13: $M = 2.11$, $SE = .30$, $t(26) = .34$, $p = .74$, $d = .07$; Adults, $M = .29$, $SE = .32$, $t(23) = 9.75$, $p < .001$, $d = 1.99$.

Non-parametric Wilcoxon signed-ranks tests of individuals' response patterns provided further insight into these data (see Table 1 for all data and statistics). Regarding conformity, all age groups most often approved of the behaviors. Regarding non-conformity, the youngest children (4 to 6) most often disapproved, older children (7 to 13) approved and disapproved at equal rates, and adults most often approved.

Negativity toward non-conformity—Next, we focused only on those children who disapproved of non-conformity and were asked how bad the behavior was (1 = *kind of bad*, 2 = *pretty bad*, 3 = *very, very bad*; *n* = 64). We had insufficient data to include the adults (only 1 adult disapproved of non-conformity), so we focused only on children (twenty-one 4- to 6-year-olds, twenty-one 7- to 9-year-olds, and twenty 10- to 13-year-olds). A univariate ANOVA with age group (3: 4–6, 7–9, 10–13) as the between-subjects variable and negativity as the dependent variable (i.e., average across non-conformity trials on which they indicated "not okay"; scores could range from 1 to 3) revealed a main effect of age group, $F(2, 59) = 9.31$, $p < .001$, $\eta_p^2 = .24$. Four-to six-year-olds ($M = 2.43$, $SE = .14$) were significantly more negative than 10- to 13-year-olds ($M = 1.56$, $SE = .14$) ($p < .001$), and 7- to 9-year-olds ($M = 2.05$, $SE = .14$) were marginally more negative than 10- to 13-year-olds ($p = .055$). Negativity did not differ significantly between the two youngest child groups ($p = .18$).

Explanations—We next turned to the explanations that participants provided after they were asked why they approved or disapproved of a given behavior (e.g., "Why is it not okay for this Hibble to speak this kind of language?"). Explanations were given for three kinds of responses: Disapproved non-conformity, approved non-conformity, and approved conformity. We did not examine explanations when participants disapproved of conformity, as this response was so rare. Because not all participants provided each type of response, we

did not statistically compare across them, but rather, examined the frequency of the four explanation types (i.e., norm-based, group-based, individual-based, similarity-based) within each response type (e.g., approved conformity). The percentages of explanation types for a given type of response were analyzed via a series of repeated measures ANOVAs in which age group was a between-subjects variable, explanation type was a within-subjects variable, and the percentage of given explanations as the dependent variable. All pairwise comparisons were Bonferroni-corrected, and unless noted, significant at the $p < .05$ level (for a similar analysis method, see Rhodes, 2014, Rutland et al., 2015). Because the explanation data violated the repeated measures assumption of sphericity, we used the Huynh-Feldt correction to adjust the degrees of freedom for the calculated F values (Field, 2011; Huynh & Feldt, 1976). The explanation data are presented in Tables 2–5.

Explanations about disapproved non-conformity: Only 1 adult disapproved of non-conformity, so adults were excluded from this analysis. Focusing on children who disapproved of non-conformity ($n = 61$), we found a significant difference in explanation type, $F(2.60, 150.89) = 25.08$, $p < .001$, $\eta_p^2 = .30$, which interacted significantly with age group, $F(5.20, 150.89) = 7.12$, $p < .001$, $\eta_p^2 = .20$. Importantly, 4- to 6-year-olds gave more norm-based and similarity-based explanations than individual-based explanations, and 7- to 9-year-olds and 10- to 13-year-olds gave more group-based explanations than any other explanation type. Thus, in disapproval of non-conformity, the youngest children primarily expressed normative rules, whereas older children focused on group membership per se.

Explanations about approved non-conformity: Focusing on participants who approved of non-conformity ($n = 64$), we found a significant main effect of age group, $F(3, 60) = 3.76$, $p = .02$, $\eta_p^2 = .16$, a significant difference in explanation type, $F(2.84, 170.45) = 10.67$, $p < .001$, $\eta_p^2 = .15$, and a significant interaction of age group and explanation type, $F(8.52, 170.45) = 2.17$, $p = .026$, $\eta_p^2 = .10$. Adults and 7- to 9-year-olds gave more individual-based explanations than any other explanation type. Thus in contrast to disapproval, which focused on group membership or norms, approval focused on individuals' wishes, desires, and choices.

Explanations about approved conformity: Focusing on participants who approved of conformity ($n = 102$), we found a main effect of age group, $F(3, 98) = 3.09$, $p = .034$, $\eta_p^2 = .08$, a significant difference in explanation type, $F(2.93, 287.33) = 17.82$, $p < .001$, $\eta_p^2 = .15$, and a significant interaction of age group and explanation, $F(5.89, 287.33) = 6.34$, $p < .001$, $\eta_p^2 = .16$. The 7- to 9-year-olds and 10- to 13-year-olds gave more group-based explanations than any other explanation type, whereas adults gave more group-based and individual-based explanations than norm-based or similarity-based explanations.

Discussion

Study 1 finds that children evaluated non-conformity more negatively than conformity, suggesting that they viewed group-based regularities as having prescriptive force. If Hibbles typically eat a certain kind of food or listen to a certain kind of music, then it is not okay for a particular Hibble to eat a different kind of food or listen to a different kind of music – indeed, it is bad. Interestingly, these patterns were strongest in the youngest participants:

Adults did not show the evaluative judgments displayed by children; and in several respects, the youngest children (4–6 years) were more negative than the older groups of children. One possibility is that the youngest children were simply more negative across the board, as they provided greater disapproval than the other age groups for conformity trials as well as non-conformity trials. However, several additional findings suggest that younger children were not simply more negative on the initial forced-choice question (okay/not okay), but rather, that they were particularly judgmental regarding non-conforming behaviors. Specifically, the youngest children were the only group that disapproved of non-conformity at levels that exceeded chance, and at the same time, disapproved of conformity at levels that were *below* chance. Furthermore, the youngest children provided stronger negativity ratings than the other groups, and in their individual response patterns, systematically disapproved of non-conformity (but approved of conformity). Finally, the youngest children explained their disapproval of non-conformity by expressing normative judgments (e.g., "Hibbles are not supposed to play games with that kind of toy"). Altogether, these data provide converging evidence to suggest that children view group regularities as prescriptive, even when the groups are novel and the behaviors are innocuous.

One methodological question concerns children's understanding of the expression "not okay", and specifically whether there are developmental changes in the semantics of this phrase that could contribute to the developmental effects that were obtained. For example, if young children have an undifferentiated concept of norms that includes both descriptive and prescriptive ideas (Bear & Knobe, 2015), their disapproval for non-conformity need not be indicative of a prescriptive judgment as it was for older children and adults (i.e., an activity could be "not okay" either because it is infrequent, or because it is wrong to do). For three reasons, however, we believe that young children's initial disapproval (i.e., evaluating something as not okay) was indeed indicative of a prescriptive judgment. First, focusing only on children who disapproved of non-conformity and were therefore presented with the follow-up negativity scale, the youngest age group showed the greatest degree of negativity, suggesting a prescriptive stance. Second, the youngest age group often appealed to prescriptive explanations when explaining their disapproval of non-conformity. Third, prior research indicates that young children initially treat ambiguous utterances (e.g., you *can* ride the bike) as deontic in meaning (e.g., you are *allowed* to ride the bike) and only later as epistemic (e.g., you are *able* to ride the bike), suggesting an early sensitivity to prescriptiveness (Modyanova et al., 2010; Papafragou, 1997).

STUDY 2

Children in Study 1 may have assumed that the novel groups were competing coalitions, given their distinct clothing patterns and behaviors, which may have licensed inferences regarding norm-appropriate behaviors within and across groups (Abrams, Rutland, Pelletier, & Ferrell, 2009; Brenick et al., 2010; Rhodes, 2012; Spielman, 2000). That is, between-group competition requires within-group coordination, which could increase prescriptiveness (e.g., if Hibbles and Glerks are working against each other, Hibbles should conform to other Hibbles), whereas between-group cooperation requires between-group coordination, which could decrease prescriptiveness (e.g., if Hibbles and Glerks are working together, there may indeed be benefits for Hibbles conforming to Glerks). Consistent with

this idea, recent work suggests that norm enforcement may be particularly likely under conditions that require within-group coordination (McAuliffe & Dunham, 2016). Alternatively, given the importance of norms for general group functioning (Chudek & Henrich, 2011), children could continue to treat group regularities as prescriptive, regardless of whether groups are competitive or cooperative with one another. Study 2 tested these alternative possibilities.

Method

Participants—Four age groups of participants were included ($N = 194$): fifty 4- to 6-year-olds (54% female, M age = 5.06 years, $SD = .77$), forty-eight 7- to 9-year-olds (54% female, M age = 7.94 years, $SD = .81$), forty-eight 10- to 13-year-olds (47% female, M age = 11.10 years, $SD = .95$), and forty-eight adults (51% female, M age = 24.88, $SD = 5.14$). Children were recruited from the same sources as those in the first study. Adults were recruited on a college campus or via Amazon's Mechanical Turk.[2] The sample was mostly white/European American (69% white/European American, 16% Asian American/Asian/Pacific Islander, 6% Multiracial, 4% black/African American, 2% Latino, and 3% other or not reported).

Materials and procedure—After being introduced to the novel groups (in a procedure that was identical to that of Study 1), participants were randomly assigned to either a competition or a cooperation condition. In both conditions, the novel groups were described as building towers out of blocks (see Rhodes, 2012 for a similar method). In the competition condition, Hibbles and Glerks were building towers against each other, there were not enough blocks for each group to build equally tall towers, and only the winning group would get a prize in the end. In the cooperation condition, Hibbles and Glerks were building a tall tower together, there were enough blocks for both groups, and both groups would get a prize in the end. To assess whether participants understood the competitive or cooperative context, they were asked whether there were enough blocks for each group. Participants who did not respond correctly were re-read the story and asked again ($n = 17$; nine at 4–6 years, seven at 7–9 years, one at 10–13 years). All participants completed these comprehension questions successfully. Adults recruited via Mechanical Turk were given an online version of the task via Qualtrics. The dependent measures were identical to those in Study 1. See Figures S2 and S3 in the online supplementary materials for sample images and vignettes.

Results

Disapproval toward conformity and non-conformity—We conducted a 4 (age group: 4–6, 7–9, 10–13, adult) × 2 (behavior: conformity, non-conformity) × 2 (condition: competition, cooperation) repeated measured ANOVA, with age group and condition as between-subjects variables, behavior as a within-subjects variable, and the frequency of not-okay evaluations as the dependent variable. There were main effects of age group, $F(3, 186) = 27.20$, $p < .001$, $\eta_p^2 = .31$, and behavior, $F(1, 558) = 312.57$, $p < .001$, $\eta_p^2 = .63$, which were qualified by a significant interaction of age group and behavior, $F(3, 186) = 14.10$, $p < .001$, $\eta_p^2 = .19$ (see Figure 2). All age groups were more disapproving of non-conformity

---

[2]Research suggests that data collected in-person or via MTurk yield comparable results (Casler, Bickel, & Hackett, 2013). An ANOVA comparing adults recruited on MTurk to those recruited in person yielded no significant differences.

than conformity ($ps < .001$). For non-conformity, all child groups were more disapproving than adults (ps .001), and 4- to 6-year-olds were more disapproving than 10- to 13-year-olds ($p = .005$), and for conformity, 4- to 6-year-olds were more disapproving than all other age groups ($ps < .001$). There were no main effects or interactions of condition.

One-sample *t*-tests compared disapproval rates against chance (i.e., 2) across age groups, behavior types, and conditions. For all age groups, disapproval was below chance for conformity (*Competition*: 4–6: $M = .57$, $SE = .10$, $t(27) = −7.07$, $p < .001$, $d = 1.34$; 7–9: $M = 0$; 10–13: $M = 0$; Adults: $M = .05$, $SE = .12$, $t(23) = −47$, $p < .001$, $d = 9.59$; *Cooperation*: 4–6: $M = .45$, $SE = .11$, $t(21) = −7.95$, $p < .001$, $d = 1.69$; 7–9: $M = .08$, $SE = .11$, $t(24) = −34.67$, $p < .001$, $d = 6.93$; 10–13: $M = .06$, $SE = .13$, $t(16) = −33$, $p < .001$, $d = 8.00$; Adult: $M = 0$). For non-conformity, the youngest age groups (4–6, 7–9) disapproved at rates above chance, 10 to 13-year-olds were at chance, and adults were below chance (*Competition*: 4–6: $M = 3.18$, $SE = .29$, $t(27) = 4.77$, $p < .001$, $d = .90$; 7–9: $M = 2.83$, $SE = .32$, $t(22) = 2.41$, $p < .025$, $d = .50$; 10–13: $M = 2.07$, $SE = .27$, $t(30) = −.43$, $p = .82$, $d = .08$; Adults: $M = 1.05$, $SE = .33$, $t(23) = −3.25$, $p = .004$, $d = .66$; *Cooperation*: 4–6: $M = 3.27$, $SE = .33$, $t(21) = 5.33$, $p < .001$, $d = 1.13$; 7–9: $M = 2.80$, $SE = .31$, $t(24) = 2.46$, $p = .022$, $d = .49$; 10–13: $M = 2.24$, $SE = .37$, $t(16) = .49$, $p < .63$, $d = .12$; Adult: $M = .70$, $SE = .29$, $t(24) = 5.02$, $p < .001$, $d = 1.00$).

Non-parametric Wilcoxon signed-ranks tests of individuals' response patterns (see Table 1) showed that regarding conformity, participants in all age groups most often approved of the behaviors. Regarding non-conformity, the youngest two age groups (4 to 6 and 7 to 9) most often disapproved, 10- to 13-year-olds approved and disapproved at equal rates, and adults most often approved.

Negativity toward non-conformity—We again focused on participants who evaluated non-conformity as "not okay" and were therefore asked how bad the behavior was ($n = 135$; forty-six 4- to 6-year-olds, thirty-nine 7- to 9-year-olds, thirty-one 10- to 13-year-olds, and nineteen adults). Because initial analyses revealed no effects of condition, we collapsed over this variable, which also provided us with enough statistical power to include adults (unlike Study 1). We conducted a 4 (age group: 4–6, 7–9, 10–13) univariate ANOVA with age group as a between-subjects variable and negativity as the dependent variable. There was a significant effect of age group, $F(3, 131) = 8.99$, $p < .001$, $\eta_p^2 = .17$. Pairwise comparisons showed that all child groups were significantly more negative than adults (4–6: $M = 2.21$, $SE = .10$; 7–9: $M = 1.92$, $SE = .11$; 10–13: $M = 1.89$, $SE = .12$; Adults: $M = 1.26$, $SE = .15$) ($ps < .02$). Child groups did not differ significantly from one another ($ps > .21$).

Explanations—Initial analyses revealed no significant condition differences, so the data were collapsed across this variable.

<u>**Explanations about disapproved non-conformity:**</u> Focusing on participants who evaluated non-conformity as not okay ($n = 112$; adults were excluded because they rarely disapproved of non-conformity, see Table 1), we found a marginally significant main effect of age group, $F(3, 109) = 2.96$, $p = .056$, $\eta_p^2 = .05$, a significant difference in explanation type, $F(2.53, 276.14) = 22.54$, $p < .001$, $\eta_p^2 = .17$, and a non-significant trend toward an age group ×

explanation type interaction, $F(5.07, 276.14) = 2.04$, $p = .072$, $\eta_p^2 = .04$. Four- to 6-year-olds gave more group-based, norm-based, and similarity-based explanations than individual-based explanations, whereas 7- to 9-year-olds and 10- to 13-year-olds gave more group-based explanations than any other explanation type. Adults gave more group-based explanations than individual-based or norm-based explanations.

<u>Explanations about approved non-conformity:</u> Focusing on participants who evaluated non-conformity as okay ($n = 110$), we found a significant main effect of age group, $F(3, 106) = 5.57$, $p = .001$, $\eta_p^2 = .14$, a significant difference in explanation type, $F(2.33, 247.12) = 55.72$, $p < .001$, $\eta_p^2 = .35$, and a significant interaction of age group and explanation, $F(6.99, 247.12) = 3.72$, $p = .001$, $\eta_p^2 = .10$. The three oldest age groups provided individual-based explanations more than any other explanation type.

<u>Explanations about approved conformity:</u> Focusing on participants who approved of conformity ($n = 191$), we found a main effect of age group, $F(3, 187) = 4.06$, $p = .008$, $\eta_p^2 = .06$, a significant difference in explanation type, $F(2.58, 482.65) = 32.46$, $p < .001$, $\eta_p^2 = .15$, and a significant interaction of age group and explanation, $F(7.74, 482.65) = 4.19$, $p < .001$, $\eta_p^2 = .06$. Four-to –six-year-olds gave more similarity-based explanations than individual-based and norm-based explanations, 7- to 9-year-olds gave more group-based explanations than any other explanation type, and 10- to 13-year-olds gave more group-based explanations than norm-based explanations. Adults gave more group-based explanations than norm-based and similarity-based explanations.

Discussion

Although we used competition and cooperation manipulations that were successful in previous research in eliciting concepts of intergroup competition and cooperation (Rhodes, 2012), and although all participants in the present study passed the manipulation checks, we obtained identical patterns in both the competition and cooperation conditions. Specifically, in both conditions, children evaluated non-conformity more negatively than conformity, and their disapproval declined with age. These data suggest that adherence to norms may be so important to human psychology—perhaps because of its role in cultural learning (Chudek & Henrich, 2011)—that descriptive regularities elicit prescriptive judgments early in development and robustly across intergroup contexts.

STUDY 3

An alternative explanation for the results of Studies 1 and 2 is that participants' judgments were not based on group regularities, but rather on any regularity, group-based or not. To test this, Study 3 presented individual regularities (behavioral patterns; e.g., "This one eats these kinds of berries") stripped of any reference to groups. We did not provide labels because they are powerful cues to group membership (Gelman & Heyman, 1999; Waxman, 2010). Study 3 included the same clothing patterns as those in Studies 1 and 2, so that non-conformity was still visually salient. We focused on 4- to 6-year-olds because they showed the highest rates of negative evaluations in Studies 1 and 2. If children's prescriptive judgments stemmed from group regularities, they should not show negative evaluations

when the focus is on individual regularities, even in the presence of visual markers of group membership.

Method

Participants—Participants included twenty-five 4- to 6-year-olds (44% female, M age = 5.00 years, SD = .82) who were recruited from the same sources as in Study 1. This sample was mostly white/European American (72% white/European American, 12% multiracial, 4% black/African American, 12% other or not reported).

Materials and procedure—The materials and procedure were identical to those in Study 1, with the exception that children were shown individuals instead of groups. The experimenter introduced children to two individuals by saying, "I'm going to tell you about these two – this one (*pointing*) and this one (*pointing*)." The two individuals were distinguished by clothing pattern only. After being introduced to the two novel individuals, participants received eight test trials that were matched to those from Study 1. On each trial, participants learned about a property for each of two individuals, and were then introduced to a third individual who either conformed or did not conform to the pattern established initially. For example, "This one listens to this kind of music (*pointing*) and this one listens to this kind of music (*pointing*). Look (*revealing the target*), this one is listening to this kind of music (*pointing*)." All dependent measures (i.e., language, games, music, food) were identical to those in Studies 1 and 2. See Figure S4 in the online supplemental materials for an example trial.

Results and Discussion

Although our primary interest was in how children's response patterns in a group-based context (Study 1) compared to an individual-based context (Study 3), we first present the data from Study 3 alone.

A paired samples t-test showed that disapproval was marginally higher for non-conformity (M = 1.48, SE = .27) than for conformity (M = .88, SE = .20), $t(24)$ = 1.23, p = .061, d = .47. One-sample t-tests showed that disapproval rates were significantly below chance (i.e., 2) for conformity, $t(24)$ = −5.32, p < .001, d = 1.06, and at chance for non-conformity, $t(24)$ = −1.80, p = .09, d = .36. These patterns were confirmed by the Wilcoxon signed-ranks tests on the individual response patterns (see Table 1). Focusing only on children who disapproved of non-conformity and were therefore asked how bad the behavior was (n = 16), children were at a mean level of 1.94 (SE = .16), suggesting that on average, they believed non-conformity to be "pretty bad". For children who disapproved of non-conformity (n = 16), group-based explanations were most common, followed by similarity based-explanations, individual-based explanations, and norm-based explanations (see Table 5). For participants who approved of non-conformity (n = 21), individual-based explanations were most common, followed by similarity-based explanations, group-based explanations, and norm-based explanations. For participants who approved of conformity (n = 25), similarity-based explanations were most common, followed by individual-based explanations, group-based explanations, and norm-based explanations. Notably, norm-based explanations were relatively rare in this study (compared to Studies 1 and 2).

Study comparison (Study 1 vs. Study 3)

Disapproval toward conformity and non-conformity: We next compared the disapproval rates and negativity ratings of 4- to 6-year-olds in Study 1 to those in Study 3. We conducted a 2 (behavior: conformity, non-conformity) × 2 (study: group-based, individual-based) repeated measures ANOVA with study as a between-subjects variable, behavior as a within-subjects variable, and the frequency of not-okay evaluations as the dependent variable. Only effects involving study are reported. There was a main effect of study, $F(1, 47) = 10.85$, $p = .002$, $\eta_p^2 = .19$, and a significant interaction of behavior and study, $F(1, 47) = 5.10$, $p = .029$, $\eta_p^2 = .10$. For conformity, disapproval rates did not differ across the two studies ($p = .52$), whereas for non-conformity, disapproval was higher in Study 1 ($p = .001$).

Negativity toward non-conformity: Next, focusing only on children who disapproved of non-conformity and were therefore asked how bad the behavior was ($n = 37$; 21 in Study 1, 16 in Study 3), we conducted a 2 (study: group-based, individual-based) univariate ANOVA with negativity as the dependent variable. The predicted significant main effect of study, $F(1, 35) = 5.17$, $p = .03$, $\eta_p^2 = .13$, showed that negativity was greater in Study 1 than in Study 3.

Explanations: We were also interested in the extent to which children's explanations differed across studies. Overall, there were no significant main or interactive effects on the basis of study. However, a planned comparison confirmed the prediction that among children who disapproved of non-conformity ($n = 38$), children in Study 1 were more likely to provide norm-based explanations than those in Study 3, $p = .047$. These comparisons did not show significant differences for the other explanation types.

In summary, as predicted, when the focus was on individual regularities (Study 3) rather than group regularities (Study 1), children 4–6 years of age less negatively evaluated non-conformity, suggesting that the negative evaluations displayed in Studies 1 and 2 stemmed from concepts of group regularities rather than regularities per se.

General Discussion

A pervasive aspect of human cognition is the tendency to use what *is* to infer what *should be* (Hume, 1738, 2000; Kalish, 2012; Nielsen & Haun, 2016). For instance, members of a group are expected to behave in specific ways (e.g., Black people are expected to speak with a certain dialect, boys are expected to wear pants), and negative judgments befall group members who violate those expectations (e.g., Black people who "talk proper", boys who wear dresses; Blakemore, 2003; Devine, 1989; Durkee & Williams, 2013; Levy et al., 1995). Moreover, the powerful ways in which regularities license prescriptive judgments emerge early in development; young children protest when someone fails to imitate irrelevant behaviors before achieving a goal, and they criticize those who fail to follow the rules of a newly learned game (Kenward, 2012; Rakoczy & Schmidt, 2013). The present studies systematically examined the conditions under which people additionally conceptualized group regularities that were novel and morally neutral (e.g., what kind of food Hibbles eat) as prescriptive (e.g., what kind of food Hibbles *should* eat). Studies 1 and 2 showed that children (ages 4 to 13) more often disapproved of non-conformity than of conformity, and that disapproval rates declined with age. Compared to older children (7–13), 4- to 6-year-

olds were more likely to disapprove of non-conformity, rate it as negative, and provide norm-based explanations when justifying their disapproval. Thus, these studies provided converging data showing that with regard to third-person, unfamiliar, and morally neutral groups, the link between what *is* and what *should be* is powerful in childhood.

The present findings have implications for research across the cognitive sciences. Conformity is important in a range of contexts (for example, as a learning strategy or as a means to gain social approval; Cialdini & Goldstein, 2004; Corriveau, Fusaro, & Harris, 2009). Our results are striking for demonstrating that people, particularly young children, negatively evaluate non-conformity even when there is no information to be learned and they have no personal stake in the given groups. These results thus suggest that an additional mechanism by which group regularities may exert influence is by rather automatically fostering an evaluative stance. An open question is whether this mechanism is found in non-human species as well as children, consistent with the proposal that there may be shared evolutionarily determined cognitive mechanisms that contribute to a cross-species preference for conformity (Claidière & Whiten, 2012).

Prescriptive judgments of non-conformity may also link to important educational and psychosocial outcomes. For instance, Black students who are accused of "acting White" (by virtue of their speech, clothing, racial identity of friends, music preferences, or academic achievement) at times disengage academically in order to avoid negative judgment, feel a decreased sense of racial identity, and show increased anxiety, depression, and emotional stress (Durkee & Williams, 2013). The present findings suggest that these judgments may be foundational in childhood, and thus potentially pervasive and difficult to modify. More generally, social groups paired with morally neutral content (e.g., "Chinese food", "Black music", "Girl toys") may foster a belief in descriptive regularities (e.g., Chinese people eat Chinese food). In turn, these regularities could license prescriptive judgments (e.g., Chinese people *should* eat Chinese food). In other words, for young children, seemingly innocuous input could generate beliefs about appropriate behavior across a wide array of content.

Critically, the observed effects were robust across different types of intergroup contexts, suggesting that children used group regularities to generate prescriptive judgments whether the novel groups were cooperating or in competition (Study 2). Of course, given the important role that coordination plays in intergroup cognition (McAuliffe & Dunham, 2016; Nielsen & Haun, 2016), it is possible that more complex coordination conditions (e.g., the groups must coordinate in order to complete a necessary goal) or behaviors that directly relate to the coordination task (e.g., someone who speaks the other group's language in order to communicate and achieve a necessary goal) would indeed moderate the extent to which descriptive regularities are used to make prescriptive judgments. Future studies would do well to test such situations, yet, for now, the present data suggest that these effects need not be moderated by intergroup contexts.

Study 3 provided an important control: When the emphasis was on individuals and not groups, children were less likely to disapprove of non-conformity, were less negative toward non-conformity, and were less likely to appeal to norm-based explanations in justifying their disapproval. This study was important in establishing that the effects in Studies 1 and 2

stemmed from group-based regularities, rather than from regularities per se. This finding also attests to how emphasizing individuality can shape group-based reasoning. A potential implication is that describing people as individuals (e.g., that person [who happens to be Black] listens to hip-hop), rather than in terms of their group membership (e.g., Black people listen to hip-hop), may reduce group-based expectations and prescriptive judgments (see also Fiske & Neuberg, 1990). An open question is which factors contributed to the differences between Study 3 and Studies 1 and 2. In Studies 1 and 2, participants were given multiple cues to group regularities: visually discernible groups (i.e., individuals were in sets of three and distinguished by clothing patterns that were shared within a group), category labels (i.e., Hibbles, Glerks), and generic statements in which labeled categories were linked to properties (e.g., Hibbles eat these kinds of berries). In Study 3, the only cue to potential cue to group regularities were clothing patterns, though note that this alone was not sufficient in licensing prescriptive judgments. Visual groups that are correlated with shared behaviors may help participants encode coalitional concepts more efficiently (Cosmides, Tooby, & Kurzban, 2003); labels imply stable, inductively rich categories (Walton & Banaji, 2004); and generic statements foster essentialism of groups (Gelman et al., 2010). Additional work is needed to test which of these factors, individually or in combination, leads children to treat descriptive regularities as prescriptively valued.

An important avenue for future research is to explore the mechanisms responsible for why prescriptive judgments declined with age. One possibility is that younger children have an undifferentiated concept of normality in which they conceptualize descriptiveness and prescriptiveness as one and the same, whereas with age, children develop the ability to differentiate between the two (Bear & Knobe, 2015). This would suggest that the descriptive-to-prescriptive link is an early emerging bias that fades as people learn that regularities per se are insufficient to prompt normative judgments. A second possibility is that older children and adults need a more complex array of factors that were intentionally removed from the present study (i.e., experiences with given groups, perceptions of social threat, more severe moral violations) to determine which features are relevant to evaluative judgments. That is, participants in all age groups may have engaged in such reasoning, but the cues elicit such reasoning may change with age. Certainly sociopolitical contexts and beliefs shape social perception (e.g., Dunham, 2011; Ho, Roberts, & Gelman, 2015), and judgments regarding non-conformity may likewise reflect such factors. A third possibility is that older children and adults hold the same descriptive-to-prescriptive intuitions as younger children, but that they suppress or override them on this explicit task. Consistent with this possibility, research on a range of stereotypes and biased attitudes (regarding race, gender, etc.) shows that whereas explicit and self-reported biases decline with age, subtle and less conscious biases persist at an implicit level (e.g., Apfelbaum, Pauker, Ambady, Sommers, & Norton, 2008; Baron & Banaji, 2006; Dunham, Baron, & Banaji, 2008; Eidsen & Coley, 2014). Future research with implicit response tasks could test whether older children and adults exhibit a subtler yet persistent descriptive-to-prescriptive link.

Additional research is also needed to more fully understand why descriptive regularities licensed children's prescriptive judgments. One possibility is that these data reflect a broader tendency to infer ought from is (see Hume, 1738/2000), though Study 3 provides evidence against a wholly unbounded version of this claim. An alternative possibility is that non-

conformity was interpreted as a moral transgression. This could explain younger children's responses; when focusing only on children who disapproved of non-conformity, younger children were most negative and most likely to appeal to rule-based explanations. As mentioned previously, older children may have relied on more sophisticated reasons to license their prescriptiveness. For instance, they may have perceived non-conformity as a violation of group loyalty, or as a violation of an immutable and innate social reality (Abrams, 2011; Diesendruck & Menahem, 2015; Tworek & Cimpian, in press). Future research should include individual difference measures (e.g., psychological essentialism) to gain further insights. Similarly, future research would also do well to directly manipulate how the groups are portrayed. Here, we gave no background information about the groups, therefore leaving open children's perceptions. Children who perceived the groups as intimately related (e.g., families, friends) may have relied on notions of group loyalty, whereas those who conceptualized the groups as task focused (e.g., sports teams) may have relied on notions on interdependence (see Lickel et al., 2000; Plötner, Over, Carpenter, & Tomasello, 2016).

The present studies provide a foundation for exploring the contexts that foster or inhibit children's tendency to treat group regularities as having prescriptive force. For instance, are children in societies that tend to value interdependence (e.g., China, Japan) more negative than children in societies that tend to value independence (e.g., the U.S., Germany, for cross-cultural insights, see Henrich, Heine, & Norenzayan, 2010; Markus & Kitayama, 1991)? Are these effects specific to the social domain, or do they hold even when reasoning about non-social entities, such as plants and artifacts (for research on domain-based category judgments, see Keil, 1989; Rhodes & Gelman, 2009)? Do evaluative judgments apply only to behaviors (e.g., eating certain foods, speaking a certain language), or do they extend to beliefs as well (e.g., holding a specific ideology; for research on children's conceptions of others' beliefs, see Heiphetz, Spelke, Harris, & Banaji, 2013; Wellman, 2014)? Do children also negatively evaluate non-conforming individuals whose group norms are themselves negative (e.g., a person who decides to help, although hurting is more typical of their group; see Cooley & Killen, 2015; Mulvey, in press; Rhodes, 2012; Smetana et al., 2012)? These important questions await future research, and promise to further reveal children's social-cognitive capacities.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Abrams D. Wherein lies children's intergroup bias? Egocentrism, social understanding, and social projection. Child Development. 2011; 82:1579–1593. [PubMed: 21790540]

Abrams D. Wherein lies children's intergroup bias? Egocentrism, social understanding, and social projection. Child Development. 2011; 82(5):1579–1593. http://doi.org/10.1111/j.1467-8624.2011.01617.x. [PubMed: 21790540]

Abrams D, Rutland A, Cameron L. The Development of Subjective Group Dynamics: Children's Judgments of Normative and Deviant In-Group and Out-Group Individuals. 2003; 74(6):1840–1856.

Abrams D, Rutland A, Pelletier J, Ferrell JM. Children's group nous: understanding and applying peer exclusion within and between groups. Child Development. 2009; 80(1):224–243. http://doi.org/10.1111/j.1467-8624.2008.01256.x. [PubMed: 19236403]

Apfelbaum EP, Pauker K, Ambady N, Sommers SR, Norton MI. Learning (not) to talk about race: When older children underperform in social categorization. Developmental Psychology. 2008; 44(5):1513–1518. http://doi.org/10.1037/a0012835. [PubMed: 18793083]

Baron AS, Banaji MR. The development of implicit attitudes: Evidence of race evaluations from ages 6 and 10 and adulthood. Psychological Science. 2006; 17(1):53–58. http://doi.org/10.1037/t03782-000. [PubMed: 16371144]

Baron AS, Dunham Y, Banaji M, Carey S. Constraints on the acquisition of social category concepts. Journal of Cognition and Development. 2014; 15:238–268. http://doi.org/10.1080/15248372.2012.742902.

Bear A, Knobe J. Folk judgments of normality: Part statistical, part evaluative. Proceedings of the Thirty-Seventh Annual Conference of the Cognitive Science Society. 2015; 184–189

Bigler RS, Jones LC, Lobliner DB. Social categorization and the formation of intergroup attitudes in children. Child Development. 1997; 68(3):530–543. http://doi.org/10.2307/1131676. [PubMed: 9249964]

Bigler RS, Liben LS. A developmental intergroup theory of social stereotypes and prejudice. Advances in Child Development and Behavior. 2006; 34:39–89. [PubMed: 17120802]

Blakemore JEO. Children's beliefs about violating gender norms: Boys shouldn't look like girls, and girls shouldn't act like boys. Sex Roles. 2003; 48(9–10):411–419. http://doi.org/10.1023/A:1023574427720.

Brenick A, Killen M, Lee-Kim J, Fox N, Leavitt L, Raviv A, Al-Smadi Y. Social Understanding in Young Israeli-Jewish, Israeli-Palestinian, Palestinian, and Jordanian Children: Moral Judgments and Stereotypes. Early Education & Development. 2010; 21(6):886–911. http://doi.org/10.1080/10409280903236598. [PubMed: 25741172]

Casler K, Bickel L, Hackett E. Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. Computers in Human Behavior. 2013; 29(6):2156–2160. http://doi.org/10.1016/j.chb.2013.05.009.

Chalik L, Rhodes M. Preschoolers Use Social Allegiances to Predict Behavior. Journal of Cognition and Development. 2014; 15(1):136–160. http://doi.org/10.1080/15248372.2012.728546.

Cialdini RB, Goldstein NJ. Social influence: compliance and conformity. Annual Review of Psychology. 2004; 55(1974):591–621. http://doi.org/10.1146/annurev.psych.55.090902.142015.

Cialdini RB, Trost MR. Social influence: Social norms, conformity and compliance. The Handbook of Social Psychology, Vols. 1 and 2. 1998 Retrieved from http://psycnet.apa.org/psycinfo/1998-07091-021.

Cimpian A, Brandone AC, Gelman Sa. Generic statements require little evidence for acceptance but have powerful implications. Cognitive Science. 2010; 34(8):1452–1482. http://doi.org/10.1111/j.1551-6709.2010.01126.x. [PubMed: 21116475]

Cimpian A, Erickson LC. Remembering kinds: New evidence that categories are privileged in children's thinking. Cognitive Psychology. 2012; 64(3):161–185. http://doi.org/10.1016/j.cogpsych.2011.11.002. [PubMed: 22197798]

Claidière N, Whiten A. Integrating the study of conformity and culture in humans and nonhuman animals. Psychological Bulletin. 2012; 138(1):126–145. http://doi.org/10.1037/a0025868. [PubMed: 22061691]

Cooley S, Killen M. Children's evaluations of resource allocation in the context of group norms. Developmental Psychology. 2015; 51(4):554–563. [PubMed: 25664833]

Corriveau KH, Fusaro M, Harris PL. Going with the flow: Preschoolers prefer nondissenters as informants. Psychological Science. 2009; 20(3):372–377. http://doi.org/10.1111/j.1467-9280.2009.02291.x. [PubMed: 19207691]

Cosmides L, Tooby J, Kurzban R. Perceptions of race. Trends in Cognitive Sciences. 2003; 7(4):173–179. http://doi.org/10.1016/S1364-6613(03)00057-3. [PubMed: 12691766]

Devine PG. Stereotypes and prejudice: Their automatic and controlled components. Journal of Personality and Social Psychology. 1989; 56(1):5–18. http://doi.org/10.1037/0022-3514.56.1.5.

Diesendruck G, Gelman SA. Domain differences in absolute judgments of category membership: evidence for an essentialist account of categorization. Psychonomic Bulletin & Review. 1999; 6(2):338–346. http://doi.org/10.3758/BF03212339. [PubMed: 12199220]

Diesendruck G, Menahem R. Essentialism promotes children's inter-ethnic bias. Frontiers in Psychology. 2015 Aug.6:1–8. http://doi.org/10.3389/fpsyg.2015.01180. [PubMed: 25688217]

Dunham Y. An angry = Outgroup effect. Journal of Experimental Social Psychology. 2011; 47(3):668–671. http://doi.org/10.1016/j.jesp.2011.01.003.

Dunham Y, Baron AS, Banaji MR. The development of implicit intergroup cognition. Trends in Cognitive Sciences. 2008; 12(7):248–253. http://doi.org/10.1016/j.tics.2008.04.006. [PubMed: 18555736]

Durkee MI, Williams JL. Accusations of Acting White: Links to Black Students' Racial Identity and Mental Health. Journal of Black Psychology. 2013; 41(1):26–48. http://doi.org/10.1177/0095798413505323.

Fordham S, Ogbu JU. Black students' school success: Coping with the "burden of "acting White".". The Urban Review. 1986; 18(3):176–206. http://doi.org/10.1007/BF01112192.

Gelman SA, Collman P, Maccoby EE. Inferring properties from categories versus inferring categories from properties: The case of gender inferring properties from categories versus inferring categories from properties: The case of gender. Child Development. 1986; 57(2):396–404. http://doi.org/10.2307/1130595.

Gelman SA, Heyman GD. Carrot-eaters and creature-believers: The effects of lexicalization on children's inferences about social categories. Psychological Science. 1999; 10(6):489–493. http://doi.org/10.1111/1467-9280.00194.

Gelman SA, Ware EA, Kleinberg F. Effects of generic language on category content and structure. Cognitive Psychology. 2010; 61(3):273–301. http://doi.org/10.1016/j.cogpsych.2010.06.001. [PubMed: 20638053]

Haun DBM, Tomasello M. Conformity to Peer Pressure in Preschool Children. Child Development. 2011; 82(6):1759–1767. http://doi.org/10.1111/j.1467-8624.2011.01666.x. [PubMed: 22023172]

Heilman ME. Gender stereotypes and workplace bias. Research in Organizational Behavior. 2012; 32:113–135. http://doi.org/10.1016/j.riob.2012.11.003.

Henrich J, Heine SJ, Norenzayan A. The weirdest people in the world? Behavioral and Brain Sciences. 2010; 33:61–135. [PubMed: 20550733]

Ho AK, Roberts SO, Gelman SA. Essentialism and racial bias jointly contribute to the categorization of multiracial individuals. Psychological Science. 2015; 26:1–7. http://doi.org/10.1177/0956797615596436.

Ho AK, Sidanius J, Cuddy AJC, Banaji MR. Status boundary enforcement and the categorization of black–white biracials. Journal of Experimental Social Psychology. 2013; 49(5):940–943. http://doi.org/10.1016/j.jesp.2013.04.010.

Kalish CW. Generalizing norms and preferences within social categories and individuals. Developmental Psychology. 2012; 48(4):1133–1143. http://doi.org/10.1037/a0026344. [PubMed: 22103303]

Kalish CW, Lawson CA. Development of social category representations: early appreciation of roles and deontic relations. Child Development. 2008; 79(3):577–593. http://doi.org/10.1111/j.1467-8624.2008.01144.x. [PubMed: 18489414]

Kenward B. Over-imitating preschoolers believe unnecessary actions are normative and enforce their performance by a third party. Journal of Experimental Child Psychology. 2012; 112(2):195–207. http://doi.org/10.1016/j.jecp.2012.02.006. [PubMed: 22436894]

Kenward B, Karlsson M, Persson J. Over-imitation is better explained by norm learning than by distorted causal learning. Proceedings. Biological Sciences/The Royal Society. 2011; 278(1709): 1239–1246. http://doi.org/10.1098/rspb.2010.1399.

Krosch AR, Amodio DM. Economic scarcity alters the perception of race. Proceedings of the National Academy of Sciences of the United States of America. 2014; 111(25):9079–9084. http://doi.org/10.1073/pnas.1404448111. [PubMed: 24927595]

Lane JD, Wellman HM, Evans EM. Children's understanding of ordinary and extraordinary minds. Child Development. 2010; 81(5):1475–1489. http://doi.org/10.1111/j.1467-8624.2010.01486.x. [PubMed: 20840235]

Lane JD, Wellman HM, Evans EM. Sociocultural input facilitates children's developing understanding of extraordinary minds. Child Development. 2012; 83(3):1007–1021. http://doi.org/10.1111/j.1467-8624.2012.01741.x. [PubMed: 22372590]

Levy GD, Taylor MG, Gelman SA. Traditional and evaluative aspects of flexibilityin gender roles, social conventions, moral rules, and physical laws. Child Development. 1995; 66(2):515–531. http://doi.org/10.1111/j.1467-8624.1995.tb00887.x. [PubMed: 7750381]

Liben LS, Bigler RS, Krogh HR. Pink and blue collar jobs: children's judgments of job status and job aspirations in relation to sex of worker. Journal of Experimental Child Psychology. 2001; 79(4): 346–363. http://doi.org/10.1006/jecp.2000.2611. [PubMed: 11511128]

Liben LS, Bigler RS, Krogh HR. Language at Work: Children's Gendered Interpretations of Occupational Titles. Child Development. 2002; 73(3):810–828. [PubMed: 12038553]

Lickel B, Hamilton DL, Wieczorkowska G, Lewis A, Sherman SJ, Uhles AN. Varieties of groups and the perception of group entitativity. Journal of Personality and Social Psychology. 2000; 78(2): 223–246. http://doi.org/10.1037/0022-3514.78.2.223. [PubMed: 10707331]

Markus HR, Kitayama S. Culture and the self: Implications for cognition, emotion, and motivation. Psychological Review. 1991; 98(2):224–253. http://doi.org/10.1037/0033-295X.98.2.224.

McAuliffe K, Dunham Y. Group bias in cooperative norm enforcement. Philosophical Transactions of the Royal Society B: Biological Sciences. 2016; 371:1–9. http://doi.org/http://dx.doi.org/10.1098/rstb.2015.0073.

McGuigan N, Robertson S. The influence of peers on the tendency of 3- and 4-year-old children to over-imitate. Journal of Experimental Child Psychology. 2015; 136:42–54. http://doi.org/10.1016/j.jecp.2015.03.004. [PubMed: 25897959]

Modyanova N, Agoos C, Kenney A, Echelbarger M, Holt A, Wexler K. Young children's interpretations of modal verbs. Language Acquisition and Development: Proceedings of Gala. 2010; (2):301–3012.

Mulvey KL. Evaluations of moral and conventional intergroup transgressions. British Journal of Developmental Psychology. 2016 http://doi.org/10.1111/bjdp.12145.

Mulvey KL, Hitti A, Rutland A, Abrams D, Killen M. When Do Children Dislike Ingroup Members? Resource Allocation from Individual and Group Perspectives. Journal of Social Issues. 2014; 70(1):29–46. http://doi.org/10.1111/josi.12045.

Nielsen M, Haun D. Why developmental psychology is incomplete without comparative and cross-cultural perspectives. Philosophical Transactions of the Royal Society B: Biological Sciences. 2016; 371(1686):1–7. http://doi.org/10.1098/rstb.2015.0069.

Nielsen M, Tomaselli K. Overimitation in Kalahari Bushman children and the origins of human cultural cognition. Psychological Science. 2010; 21(5):729–736. http://doi.org/10.1177/0956797610368808. [PubMed: 20483853]

Olson KR, Dweck CS. A blueprint for social cognitive development. Perspectives on Psychological Science. 2008; 3(3):193–202. [PubMed: 26158935]

Papafragou A. Modality in language development: a reconsideration of the evidence. University College London Working Papers in Linguistics. 1997; 9:1–31. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.11.3373&rep=rep1&type=pdf.
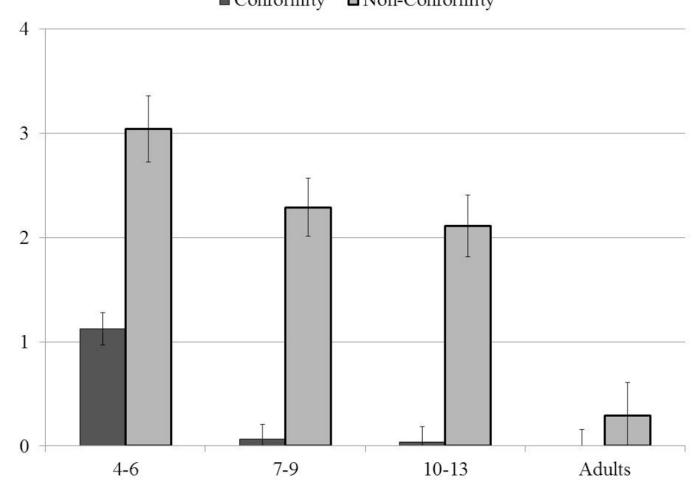
Plötner M, Over H, Carpenter M, Tomasello M. What is a group? Young children's perceptions of different types of groups and group entitavity. Plos One. 2016; 11(3):1–14. http://doi.org/10.1371/journal.pone.0152001.

Quintana SM. Children's developmental understanding of ethnicity and race. Applied and Preventive Psychology. 1998; 7(1):27–45. http://doi.org/10.1016/S0962-1849(98)80020-6.

Rakoczy H, Schmidt MFH. The early ontogeny of social norms. Child Development Perspectives. 2013; 7(1):17–21. http://doi.org/10.1111/cdep.12010.

Rhodes M. Naïve theories of social groups. Child Development. 2012; 83(6):1900–1916. http://doi.org/10.1111/j.1467-8624.2012.01835.x. [PubMed: 22906078]

Rhodes M. Children's explanations as a window into their intuitive theories of the social world. Cognitive Science. 2014; 38(8):1687–1697. http://doi.org/10.1111/cogs.12129. [PubMed: 25052813]

Rhodes M, Brickman D. The influence of competition on children's social categories. Journal of Cognition and Development. 2011; 12:194–221. http://doi.org/10.1080/15248372.2010.535230.

Rhodes M, Chalik L. Social categories as markers of intrinsic interpersonal obligations. Psychological Science. 2013; 24(6):999–1006. http://doi.org/10.1177/0956797612466267. [PubMed: 23613213]

Rhodes M, Gelman SA. A developmental examination of the conceptual structure of animal, artifact, and human social categories across two cultural contexts. Cognitive Psychology. 2009; 59(3):244–274. http://doi.org/10.1016/j.cogpsych.2009.05.001. [PubMed: 19524886]

Rhodes M, Leslie S-J, Tworek CM. Cultural transmission of social essentialism. Proceedings of the National Academy of Sciences. 2012; 109(34):13526–13531. http://doi.org/10.1073/pnas.1208951109.

Roberts SO, Gelman Sa. Do Children See in Black and White? Children's and Adults' Categorizations of Multiracial Individuals. Child Development. 2015a; 00(0):n/a–n/a. http://doi.org/10.1111/cdev.12410.

Roberts SO, Gelman SA. Do children see in Black and White? Children's and adults' categorizations of multiracial individuals. Child Development. 2015b; 86(6):1830–1847. http://doi.org/10.1111/cdev.12410. [PubMed: 26315349]

Rodeheffer CD, Hill SE, Lord CG. Does this recession make me look Black? The effect of resource scarcity on the categorization of biracial faces. Psychological Science. 2012; 23(12):1–3. http://doi.org/10.1177/0956797612450892.

Rothbart M, Taylor M. Category labels and social reality: Do we view social categories as natural kinds? 1992

Rutland A, Hitti A, Mulvey KL, Abrams D, Killen M. When does the in-group like the out-group? Bias among children as a function of group norms. Psychological Science. 2015; 26(6):834–842. http://doi.org/10.1177/0956797615572758. [PubMed: 25888686]

Schillaci RS, Kelemen D. Children's conformity when acquiring novel conventions: The case of artifacts. Journal of Cognition and Development. 2014; 15(4):569–583. http://doi.org/10.1080/15248372.2013.784973.

Schmidt MFH, Rakoczy H, Tomasello M. Young children attribute normativity to novel actions without pedagogy or normative language. Developmental Science. 2011; 14(3):530–539. http://doi.org/10.1111/j.1467-7687.2010.01000.x. [PubMed: 21477192]

Shutts K, Pemberton Roben CK, Spelke ES. Children's use of social categories in thinking about people and social relationships. Journal of Cognition and Development. 2013; 14(1):35–62. http://doi.org/10.1080/15248372.2011.638686. [PubMed: 23646000]

Smetana JG. Preschool children's conceptions of moral and social rules. Child Development. 1981; 52(4):1333–1336. http://doi.org/10.2307/1129527.

Smetana JG, Jambon M, Conry-Murray C, Sturge-Apple ML. Reciprocal associations between young children's developing moral judgments and theory of mind. Developmental Psychology. 2012; 48(4):1144–1155. http://doi.org/10.1037/a0025891. [PubMed: 22004340]

Spielman DA. Young children, minimal groups, and dichotomous categorization. Personality & Social Psychology Bulletin. 2000; 26:1433–1441.

Thomsen L, Green EGT, Sidanius J. We will hunt them down: How social dominance orientation and right-wing authoritarianism fuel ethnic persecution of immigrants in fundamentally different ways.

Journal of Experimental Social Psychology. 2008; 44(6):1455–1464. http://doi.org/10.1016/j.jesp.2008.06.011.

Tworek CM, Cimpian A. Why do people tend to infer ought from is? The role of biases in explanation. Psychological Science. 2016

Waxman SR. Names will never hurt me? Naming and the development of racial and gender categories in preschool-aged children. European Journal of Social Psychology. 2010; 40:593–610. http://doi.org/10.1002/ejsp.

Figure 1.
Study 1. Mean frequency of disapproval of conformity and non-conformity across age
groups. Scores could range from 0 to 4. Error bars depict standard errors.
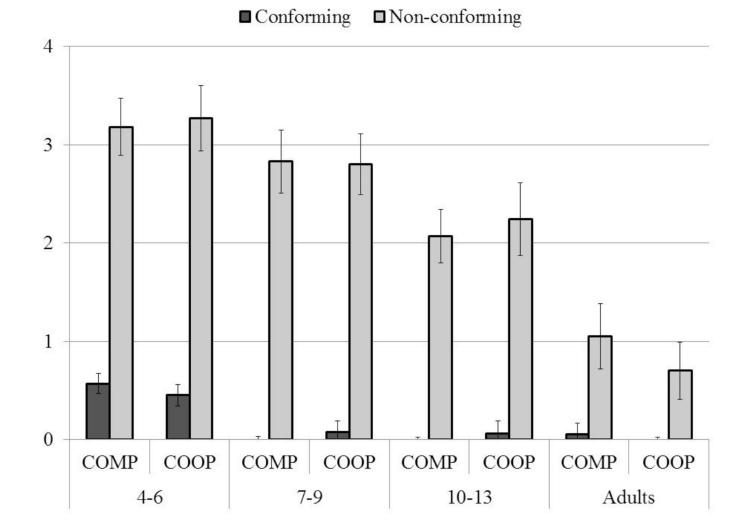
Figure 2.
Study 2. Mean frequency of disapproval of conformity and non-conformity across age groups and conditions. Scores could range from 0 to 4.

Table 1

Non-parametric tests across studies, age group, and behavior, indicating how many participants most often approved (OK), most often disapproved (Not OK), or approved and disapproved equally (Tie).

Study	Age	Behavior	OK	Not OK	Tie	Z	p	r
1	4–6	Non-conformity	5	18	1	2.86	.004	.58
		Conformity	16	5	3	−2.39	.017	.49
	7–9	Non-conformity	13	16	2	1.10	.27	.20
		Conformity	30	0	1	−5.48	<.001	.98
	10–13	Non-conformity	12	12	3	.54	.59	.10
		Conformity	27	0	0	−5.11	<.001	.98
	Adults	Non-conformity	23	1	0	−4.23	<.001	.86
		Conformity	24	0	0	−4.89	<.001	1.00
2	4–6	Non-conformity	5	40	5	4.75	<.001	.67
		Conformity	41	3	6	−5.88	<.001	.83
	7–9	Non-conformity	12	32	4	3.14	.002	.45
		Conformity	48	0	0	−4.80	<.001	.69
	10–13	Non-conformity	22	25	1	.49	.62	.08
		Conformity	48	0	0	−6.86	<.001	.99
	Adults	Non-conformity	38	8	2	−4.51	.001	.65
		Conformity	48	0	0	−6.86	<.001	.99
3	4–6	Non-conformity	13	5	7	−1.58	.115	.02
		Conformity	17	2	6	−3.69	.001	.74

Note. The data for Study 2 are collapsed across conditions (i.e., Competition, Cooperation) as each condition produced identical effects.

Table 2

Study 1. Percentage of explanation types for each behavior, across evaluation types and age groups.

Behavior	Evaluation	Age	N	Percentage of Explanation Types M(SE)			
				Norm	Group	Individual	Similarity
Non-Conformity	Not Okay	4–6	21	32.9(8.9)	23.1(10.1)	1.6(3.7)	35(7.2)
		7–9	21	19(7.2)	70.8(9)	1(3.8)	17.7(6.3)
		10–13	19	11.9(9)	48.3(11.2)	4.2(4.7)	2.8(7.9)
		Adult	4	6.3(15.6)	62.5(20.1)	25(8.4)	0
Non-conformity	Okay	4–6	9	15.8(11.4)	0	13.3(13.6)	19.4(10.1)
		7–9	15	27.5(8.9)	22.9(8.6)	67.1(10.7)	5(6.5)
		10–13	17	30.6(8.2)	20.4(8)	33(9.9)	10(6.1)
		Adult	23	21.6(6.8)	24.3(6.6)	59(8.2)	1.3(5)
Conformity	Okay	4–6	20	16.7(6.6)	7.8(7.8)	9.7(6.1)	29.2(7.3)
		7–9	31	4.7(5.2)	65.5(6.1)	17.4(5)	14.4(5.1)
		10–13	27	11.8(5.9)	50.2(7)	16(5.6)	9.5(5.8)
		Adult	24	9.8(5.9)	40.4(7)	34.8(5.7)	3.2(5.8)

Note. Scores represent each type of explanation out of the total number of trials. Across studies, individual explanations could have been coded as of more than one type, and explanations that did not fit any of the coded types are not reported (this explains why the percentages can add to more than or less than 100). Data for disapproved conformity are not presented because this response was rarely given.

Table 3

Study 2 (Competition Condition). Percentage of explanation types for each behavior, across evaluation types and age groups.

Behavior	Evaluation	Age	N	Norm	Group	Individual	Similarity
						Percentage of Explanation Types M(SE)	
Non-Conformity	Not Okay	4–6	25	31.8(6)	35.8(7.4)	0	13.3(4.8)
		7–9	17	26.3(7.3)	46.3(9.1)	4.3(3.9)	17.4(5.9)
		10–13	20	9.8(6.5)	37.2(8)	7(3.4)	8.9(5.2)
		Adult	11	2.4(9)	45.4(11.2)	7.9(4.8)	10.9(7.3)
Non-conformity	Okay	4–6	10	2.8(5.2)	2.8(8.9)	14.2(10.8)	28.1(3.4)
		7–9	9	0	11.7(9.3)	73.8(11.4)	3.1(3.5)
		10–13	19	7.2(3.7)	18.2(6.4)	67.2(7.7)	1.6(2.4)
		Adult	21	8.3(3.5)	16.8(6.1)	51.2(7.3)	2.8(2.3)
Conformity	Okay	4–6	27	7.9(3.6)	33.5(17.1)	12.3(5.9)	23.6(5.8)
		7–9	21	10.4(4.1)	50.3(8.1)	20.5(6.7)	23.6(6.5)
		10–13	31	4.7(3.4)	40.8(6.7)	20.3(5.5)	25.1(5.3)
		Adult	23	14.1(3.9)	44.9(7.7)	30.7(6.3)	8.8(6.1)

Table 4

Study 2 (Cooperation Condition). Percentage of explanation types for each behavior, across evaluation types and age groups.

Behavior	Evaluation	Age	N	Percentage of Explanation Types M(SE)			
				Norm	Group	Individual	Similarity
Non-Conformity	Not Okay	4–6	21	14.0(6.3)	19.0(7.9)	8.6(3.4)	29.4(5.1)
		7–9	20	14.5(6.5)	43.4(8.0)	9.4(3.4)	19.3(5.2)
		10–13	10	13.1(9.9)	51.2(12.3)	2.4(5.3)	11.9(8.0)
		Adult	9	5.0(9.6)	35.0(12.0)	5.0(5.1)	16.5(7.8)
Non-conformity	Okay	4–6	10	0	13.3(8.7)	23.3(10.5)	5.0(5.1)
		7–9	11	3.6(5.4)	8.9(9.4)	35.9(11.5)	3.6(3.5)
		10–13	7	10.4(6.1)	18.7(10.6)	52.1(12.7)	7.3(4.0)
		Adult	27	6.9(3.1)	34.1(5.4)	47.1(6.5)	3.5(2.0)
Conformity	Okay	4–6	22	5.6(4.0)	17.1(7.9)	8.5(6.5)	42.8(6.3)
		7–9	25	6.1(3.7)	50.0(7.4)	10.2(6.1)	22.4(5.9)
		10–13	17	10.2(4.7)	37.1(9.4)	21.6(7.7)	20.5(7.5)
		Adult	29	3.2(3.5)	49.9(7)	28.6(5.7)	6.8(5.5)

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 5

Study 3. Percentage of explanation types for each behavior, across each evaluation type.

Behavior	Evaluation	Age	N	Percentage of Explanation Types M(SE)			
				Norm	Group	Individual	Similarity
Non-conformity	Not Okay	4–6	16	1.0(6.6)	13.0(4.8)	28.7(6.0)	29.3(8.2)
Non-conformity	Okay	4–6	21	2.4(4.4)	3.6(3)	42.9(8.6)	22.2(7.4)
Conformity	Okay	4–6	25	8.8(8.3)	33.8(9.3)	9.8(3.8)	27.9(9)