

So It Is, So It Shall Be: Group Regularities and Prescriptive Judgments

by

Steven O. Roberts

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Psychology)
in the University of Michigan
2017

Doctoral Committee:

Professor Susan A. Gelman, Chair
Associate Research Scientist, E. Margaret Evans
Assistant Professor Arnold K. Ho
Professor Stephanie J. Rowley
Professor Henry M. Wellman

Steven O. Roberts

sothello@umich.edu

ORCID iD: 0000-0002-9308-4295

To my grandmother, Lucille Humphrey, for forcing me to do better.

ACKNOWLEDGMENTS

This dissertation was filtered through the brilliant minds of Susan A. Gelman, who served as my primary adviser and committee chair, and Arnold K. Ho, who served as my secondary adviser and co-chair. Without their mentorship and patience, the following pages would have remained empty. I also thank my other committee members: Henry H. Wellman, Stephanie J. Rowley, and E. Margaret Evans, for their critical feedback, and more importantly, for their investment in me.

This research would not have been possible without the many research assistants who helped collect, enter, and code data: Madeline Adelman, Paulina Bromberg, Justine Ezelle, Elizabeth Garcia, Melissa Granville, Ines Hadziegric, Ethan Jannott, Laura Kanji, Hyeongjun “Jack” Kim, Kerrie Leonard, Shira Lossos, Kevin Ma, Antonio Malkoun, Pragya Mathur, Sarah Romberg, Rebecca Strauss, Nikea Turner, Abigail Tzau, Vanessa “Ji Yoon” Lee, and Tiffany Valencia. I thank them for their hard work and dedication to this research. I am also grateful for Cai Guo, who made Chapter 4 possible.

I would not have had the strength to write were it not for my kind, compassionate, and intelligent friends in Ann Arbor. Paige Safyer showed me how to pursue, establish, and maintain meaningful friendships, and she made me a better human. Francisco Velasquez mended my broken heart and inspired me to dance. Selin Gulgoz showed me how to value, respect, and appreciate those around me, and subsequently made me a better colleague and friend. Soraya Giaccardi forced me to take life less seriously, and she held my hand as I underwent surgery. Morgan Jerald supported me although she had no reason to, and I cherished our many

discussions about research, race, gender, and family. Maria Arredondo laughed at my jokes even though they were not always funny, and she [supposedly] introduced me to whiskey. Patty Kuo and Robert James Russell showed me what a beautiful partnership looked like (as did Francisco and Selin), and they fed me when I could not eat. Joseph Bayer made my birthdays spectacular, Josh Wondra taught me about empathy (many, many times), and Neil Lewis Jr. and I honored our ancestors. Babe Kawaii-Bogue helped maintain my unapologetic self, and along with George Smith, Adrian Gale and Nkemka Anyiwo, she kept me healthy. Lauren Reed and Kimberly Brink showed me the value and beauty in cinema, and Berenice Castillo reminded me that everything can be hilarious. Omari W. Keeles taught me about race and class in America, and introduced me to the beauty of orchestra. Casta Guillaume and Alvin Thomas fueled my competitiveness and made sure my head never got too big. Amber Williams and Myles Durkee sharpened my thinking and became valued friends and collaborators, and Rosanne Jocson and Hannah Giasson (along with Paige, Soraya, and Jaron), brought great music, food, and laughter into my life. Together, these wonderful humans made my years in Ann Arbor unimaginably beautiful.

I would also like to thank several remarkable women to whom I owe my career. Kimberly Russell inspired me to leave Babies “R” Us and attend the Borough of Manhattan Community College (BMCC). Allana Hankey-Thomas inspired me to leave BMCC and attend NYU. Rose K. Vukovic inspired me to leave NYU and attend the University of Michigan (and to avoid the U.S. Army). Susan A. Gelman inspired me to leave Michigan and become an Assistant Professor of Psychology at Stanford. Also, Stephanie J. Rowley made sure that Michigan Psychology felt like home, Jacqueline Mattis inspired me with every word she spoke and every hug she gave, and Enrique Neblett Jr. showed me how to think BIG. L. Monique Ward

broadened my perspective on the academy, as well as my understanding of friendship. I am not religious, though I am convinced that these women (and Enrique) were angels sent from someone, somewhere, to see me along my way.

I thank my family, which is the root of the characteristics and values that made this dissertation inevitable. Sabine Roberts, my mother, gave me compassion, love for others, and an understanding of what is truly meaningful in life. Troy Roberts, my father, gave me pride, a sharp and somewhat inappropriate sense of humor, and the drive to lead and not follow. Sabrina, Alicia, Patrick, Celina, and Luca, my siblings, inspired me, as the oldest, to be a role model worth following. I look forward to seeing them outshine me in all aspects and I hope they don't forget me when they do. Jimmy Anderson and Eon Smith, my uncles, served as intellectual role models, and they helped me become a man. Justin Anderson and David Gutshall, my cousins, and Nikolai Dubow, my oldest friend, motivated me from a distance and always kept me grounded. Barbara Roberts, my step-mother, was a blessing who helped repair a family that had nearly fallen apart. Othello Hines Jr., my biological father, inspired my research interest in social biases, and the horrible consequences they have. I often found comfort in the thought that this research, somehow, might do his murder justice. Adelheid Wiltrud Heinzelmännchen, meine Oma, war meine erste Liebe und die Hauptfigur in meine erste Erinnerung. Wann immer es mir schlecht ging, dachte ich an sie, und an alle Kekse die Sie für uns gebacken hat.

Lastly, Lucille Humphrey, my grandmother, saved my life. When I was lost, afraid, and hopeless, she gave me the vision to see clearly, the confidence to feel strong, and the inspiration to dream. Like an alchemist, she transformed metal into gold. This dissertation is dedicated to her.

TABLE OF CONTENTS

DEDICATION	ii
ACKNOWLEDGEMENTS	iii
LIST OF TABLES	vii
LIST OF FIGURES	viii
LIST OF APPENDICES	ix
ABSTRACT	x
CHAPTER	
I. Introduction	1
II. So it is, so it shall be: Group regularities license children’s prescriptive judgments	5
III. Minimal perceptual and linguistic cues foster children’s prescriptive judgments	41
IV. Children’s descriptive-to-prescriptive tendency replicates (and varies) cross-culturally	66
V. General Discussion	89
APPENDICES	95
REFERENCES	100

LIST OF TABLES

1.	Description of coding scheme (Chapters II, III, & IV)	29
2.	Non-parametric data (Chapter II, Study I)	30
3.	Explanation data (Chapter II, Study I)	31
4.	Explanation data (Chapter II, Study II, Competition Condition)	32
5.	Explanation data (Chapter II, Study II, Cooperation Condition)	33
6.	Explanation data (Chapter II, Study III)	34
7.	Disapproval data (Chapter III)	59
8.	Non-parametric data (Chapter III)	60
9.	Explanation data (Chapter III, younger children)	61
10.	Explanation data (Chapter III, older children)	62
11.	Parent survey (Chapter IV)	84
12.	Disapproval data (Chapter IV)	85
13.	Non-parametric data (Chapter IV)	86
14.	Explanation data (Chapter IV)	87

LIST OF FIGURES

1.	Example trial (Chapter II, Study I)	35
2.	Disapproval data (Chapter II, Study I)	36
3.	Competition vignette (Chapter II, Study II)	37
4.	Cooperation vignette (Chapter II, Study II)	38
5.	Disapproval data (Chapter II, Study II)	39
6.	Example trial (Chapter II, Study III)	40
7.	Example trial (Chapter III, Group Presence condition)	63
8.	Example trial (Chapter III, Label, Generic, and Control conditions)	64
9.	Disapproval data (Chapter III)	65
10.	Disapproval data (Chapter IV, cultural comparison)	88

LIST OF APPENDICES

A.	Item effects (Chapter I, Study I)	95
B.	Chinese and English Scripts	99

ABSTRACT

Children negatively evaluate those who fail to conform to group norms (e.g., doctors who harm people, boys who wear lipstick; Kalish, 2012), yet to what extent do groups per se evoke a prescriptive stance? This was unaddressed in previous research, which provided additional cues that may have encouraged a prescriptive stance (e.g., moral principles, membership in one of the groups, or cultural input regarding non-conformity). In this dissertation, I tested whether children interpreted innocuous behaviors of novel groups (harmless characteristics shared by individuals within an unfamiliar group) as prescriptive (characteristics that individuals *should* do). Children ages 4-13 and adults were introduced to two novel groups: Hibbles and Glerks, who engaged in innocuous behaviors (i.e., the kind of music they listened to, berries they ate, games they played, and language they spoke). They were then shown conforming and non-conforming individuals and were assessed on the extent to which they approved or disapproved of their actions (measured through evaluations, negativity ratings, and open-ended explanations).

In Chapter II, I report three studies finding that children disapproved of non-conformity and justified their disapproval through norm-based reasoning (e.g., “Hibbles are not *supposed* to do that”; Study 1). These effects replicated across competitive and cooperative intergroup contexts (Study 2) and stemmed from reasoning about group norms rather than norms applied to individuals (Study 3). In Chapter III, to more precisely understand what information children used to detect group norms, a new group of children was randomly distributed across four conditions that manipulated how group norms were presented: group labels, generic statements, visual groups, or control. Children disapproved of non-conformity in all but the control

condition. Because U.S. society tends to value independence over interdependence (Markus & Kitayama, 1991), U.S. children may come to perceive non-conformity as an expression of individuality, and thus evaluate it less negatively than children from societies that tend to value interdependence. I tested this in Chapter IV by assessing children and adults recruited in Jianshi, China (population ~ 500,000). Paralleling U.S. children, Jianshi children disapproved of non-conformity and their rates of disapproval declined with age. In contrast to U.S. children, however, they remained relatively more disapproving at an older age.

Collectively, these findings demonstrate that children used group norms to generate prescriptive judgments, and that this tendency emerged 1) when reasoning about innocuous behaviors in novel groups, 2) with only minimal perceptual and linguistic input, and 3) across cultures, though variable in the rate at which it declined across development. I will discuss the implications for how readily children engage in group-based norm enforcement and stereotyping, as well as the theoretical and practical significance of the present findings, and will detail concrete directions for future research.

CHAPTER I

Introduction

Social groups consist of a collection of individuals believed to share certain properties. For example, physicians are expected to heal people, boys are expected to not wear lipstick, and Ph.D. students are expected to write dissertations (Devine, 1989; Kalish, 2012). Importantly, social groups are often believed to consist of individuals who *should* share certain properties: Physicians who don't help others, boys who wear lipstick, or Ph.D. students who do not work on their dissertations may be evaluated negatively (Blakemore, 2003; Heilman, 2012; Levy, Taylor, & Gelman, 1995). Thus, once children (and adults) believe that a social group *is* characterized by a property (i.e., descriptive group regularity), they tend to believe that individual group members *should be* characterized by that property (i.e., prescriptive group norm). This “descriptive-to-prescriptive tendency” is at the heart of the current work. To what extent do children use descriptive group regularities to generate prescriptive judgments?

Previous research in developmental, cognitive, and social psychology documents that factors that contribute to children's negative evaluation of non-conformity include familiarity with groups (e.g., gendered language and segregation tell children about how boys and girls should be; Bigler & Liben, 2006), in-group biases (e.g., children believe that in-group members who violate group norms are threatening to the group's cohesion and functioning; Rutland, Hitti, Mulvey, Abrams, & Killen, 2015), moral concerns (e.g., it is unacceptable to harm other; Smetana, 1981), and coalitional concepts (e.g., children often believe that when social groups are

in competition, individual group members should conform to their own group; Rhodes, 2012). The primary goal of my dissertation work is to examine whether children interpret regularities that are purely descriptive as prescriptive, in the absence of these factors, thereby revealing a more intuitive mechanism through which children enforce group norms and stereotypes.

I predicted that children's disapproval of group norm non-conformity would be rooted in their reasoning about norms. That is, preschool-aged children are quick to recognize, adopt, create, and enforce norms across several social domains (Rakoczy & Schmidt, 2013). For example, they protest when someone violates the rules of a newly learned game (Rakoczy, Hamann, Warneken, & Tomasello, 2010) and they create and teach their own rules when no rules are provided (Göckeritz, Schmidt, & Tomasello, 2014). When they observe a tool used in an unnecessary way to accomplish a goal, they use the tool in the same way and criticize those who do not, believing that the action reflects how the tool *should be* used (Kenward, 2012; Nielsen & Tomaselli, 2010). Children also interpret people's actions as normative even in the absence of pedagogy or normative language (Schmidt, Rakoczy, & Tomasello, 2011), and even after seeing an action occur only once (Schmidt, Butler, Heinz, & Tomasello, 2016). With this research in mind, I predict that children interpret social groups as normative, thereby taking descriptive group regularities and using them to generate prescriptive judgments.

This research has important implications for a variety of disciplines, ranging from comparative biology to norm enforcement, by providing insight into the cognitive mechanisms that contribute to stereotyping and norm enforcement (Claidière & Whiten, 2012; Devine, 1989; McAuliffe & Dunham, 2016). For example, it reveals how readily children use group norms to evaluate and judge others, and provides insight into how readily they maintain and enforce the correctness of those norms.

The present studies

To test the extent to which children used descriptive group regularities to make prescriptive judgments, I present them with two novel and third-party groups: Hibbles and Glerks, engaged in innocuous behaviors that contrast with one another on four dimensions (i.e., the kind of music they listen to, language they speak, food they eat, and games they play). Participants are then introduced to conforming and non-conforming individuals (e.g., a Hibble who listens to music more typical of Glerks). I measure participants' *disapproval* of the behaviors, degree of *negativity* toward the behaviors they disapprove of, and open-ended *explanations* justifying their evaluations. I 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). I focus on children from 4-13 years of age, because during this period there are important changes in children's school-based and peer-based experiences and their social reasoning (e.g., with regard to categorization, identity, and stereotyping), and as a comparison to related work also focused on this age period (Bigler & Liben, 1997; Quintana, 1998; Rhodes & Gelman, 2009; Roberts & Gelman, 2015). Adults provide an important comparison group to assess mature patterns of reasoning. Including preschool-aged children is especially useful, as they have fewer social experiences than their older peers. Thus, including a wide range of age groups is useful for testing the extent to which children's descriptive-to-prescriptive tendency is early emerging and intuitive, or developed with age and experience.

In Chapter II, I report data on the strength of children's descriptive-to-prescriptive tendency, as well as the extent to which it declined with age. I also report data on whether this tendency replicates across intergroup contexts (i.e., whether the novel group is shown to be in

competition or in cooperation), and whether it is reduced by focusing children on individuals rather than groups. These data are published in *Cognitive Science* (Roberts, Gelman, & Ho, 2017a). In Chapter III, I report data on how visual and linguistic cues contribute to this tendency. These data are published in the *Journal of Experimental Child Psychology* (Roberts, Ho, & Gelman, 2017b). The data reported in Chapter IV address whether children's descriptive-to-prescriptive tendency replicates internationally. These data are published in the *Journal of Experimental Child Psychology* (Roberts, Guo, Ho, & Gelman, 2017c). Taken together, the research presented here provides strong empirical support for an early emerging descriptive-to-prescriptive tendency that emerges 1) when reasoning about innocuous behaviors of novel groups, 2) with groups cued by minimal visual and linguistic cues, and, 3) across two distinct cultural contexts.

CHAPTER II

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

Abstract

The research reported in this chapter 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, have important implications for understanding the development of stereotyping and norm enforcement, and lay the foundation for the subsequent studies.

Introduction

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, Bigler, & Krogh, 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 et al., 2015). Others studies included morally laden behaviors (e.g., harming others) that have inherent meaning to children as well as adults (e.g., Mulvey, 2016; Smetana, Conry-Murry, & Sturge-Apple, 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.

In the present research, 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, 2017a). 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

¹As 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.

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 non-conforming individual and the other half depicted a conforming individual (for Figure 1 for an example trial).

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). See Table 1 for the coding scheme.

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. 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) x 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 2.

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 2 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 3-6.

Explanations about disapproved non-conformity

Only 4 adults 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, younger 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, 2016), 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

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.

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 3 and 4 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) x 2 (behavior: conformity, non-conformity) x 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 5). All age groups were more disapproving of non-conformity than conformity ($ps < .001$). For non-conformity, all child groups were more disapproving than adults ($ps \leq .001$), 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 2) 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.

Explanations about disapproved non-conformity

Focusing on participants who evaluated non-conformity as not okay ($n = 112$; adults were excluded because they rarely disapproved of non-conformity, see Table 2), 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 x 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.

Explanations about approved non-conformity

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.

Explanations about approved conformity

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 6.

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 2). 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 6). 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) x 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, 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-party, 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 clearly demonstrate that 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).

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). I turned to these questions in Chapter 3.

Table 1.

Description of coding scheme (Chapters II, III, & IV)

Code	Description	Examples
Norm	Explicitly mentions that there is a rule or an obligation that the individual must adhere to.	“They are supposed to play with that kind of toy.” “They have to listen to that kind of music.” “They aren’t allowed to do that.”
Group	References the groups or the category labels.	“Because that's what Glerks do.” “Because it's a Hibble.” “Because that’s what the rest of the group is doing.”
Individual	References mental states, including thoughts, emotions, motivations, traits. States that the behavior is about personal choice rather than group membership.	“They can do whatever they want.” “They can do that if they like it or want to.” “Different people like different things.”
Similarity	Mentions similar or dissimilar physical appearances.	“Because they are orange.” “Because it looks like that one.” “Green goes with green.”
Other	Gives an explanation that does not fit the others codes.	“I don’t like them.” “It is weird.”

Table 2.

Non-parametric data (Chapter II, Study 1). Data are 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 3.

Explanation data (Chapter II, Study 1). Percentage of explanation types for each behavior, across evaluation types and age groups.

Percentage of Explanation Types							
M(SE)							
Behavior	Evaluation	Age	N	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 4.

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

Percentage of Explanation Types							
M(SE)							
Behavior	Evaluation	Age	N	Norm	Group	Individual	Similarity
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 5.

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

Percentage of Explanation Types							
M(SE)							
Behavior	Evaluation	Age	N	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)

Table 6.

Explanation data (Chapter II, Study 3). Percentage of explanation types for each behavior, across each evaluation type.

Percentage of Explanation Types							
M(SE)							
Behavior	Evaluation	Age	N	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)

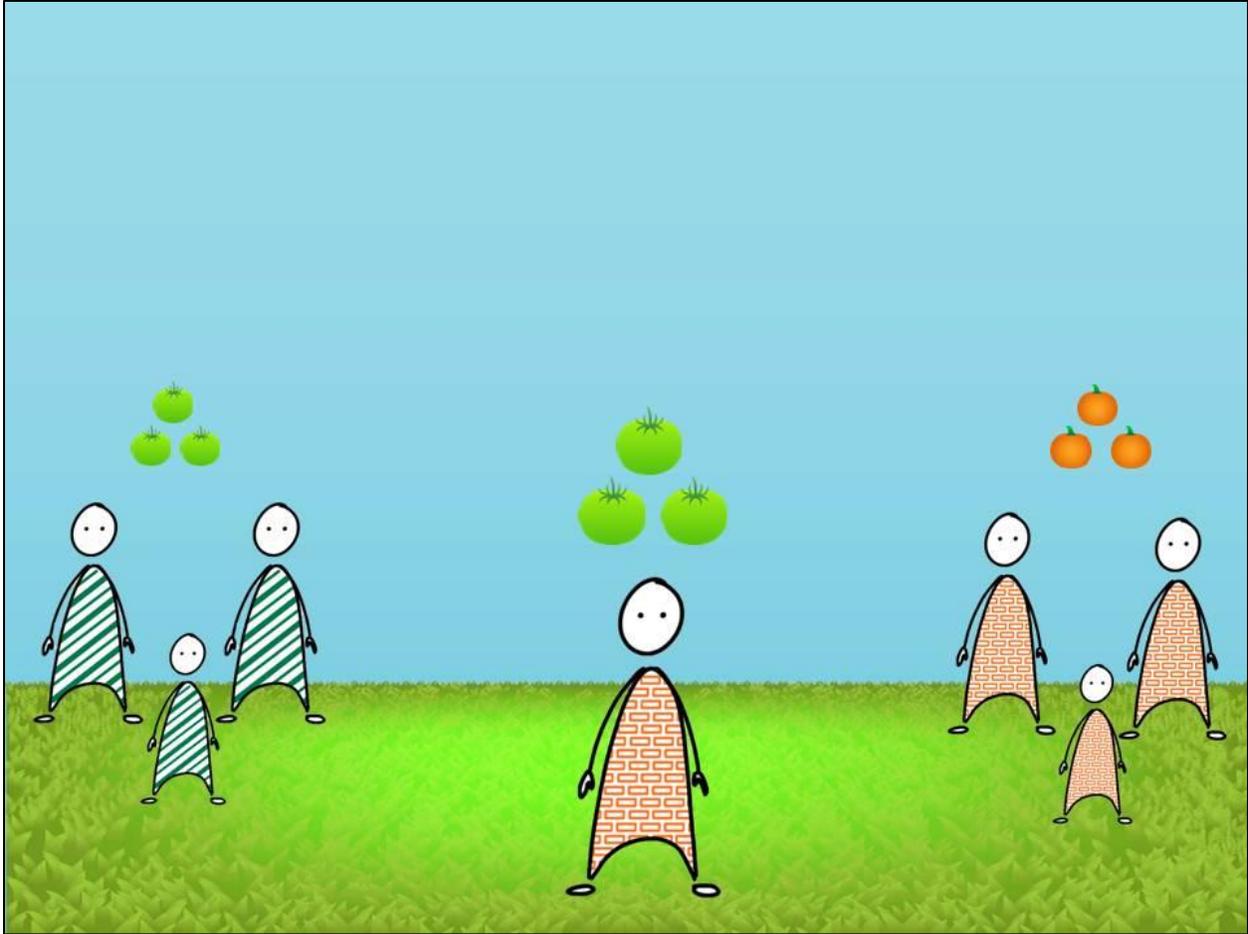


Figure 1. Example trial (Chapter II, Study 1).

“Hibbles (pointing) eat these kind of berries and Glerks (pointing) eat these kinds of berries. Look, this Glerk (pointing to target) is eating these kind of berries.

Evaluation: “Is it okay or not okay for this Glerk to eat these kind of berries?”

Negativity: [if participant evaluated this as not okay] “Is it a little bad, pretty bad, or very, very bad, for this Hibble to eat these kind of berries?”

Explanation: “Why is it [okay or not okay] for this Glerk to eat these kind of berries?”

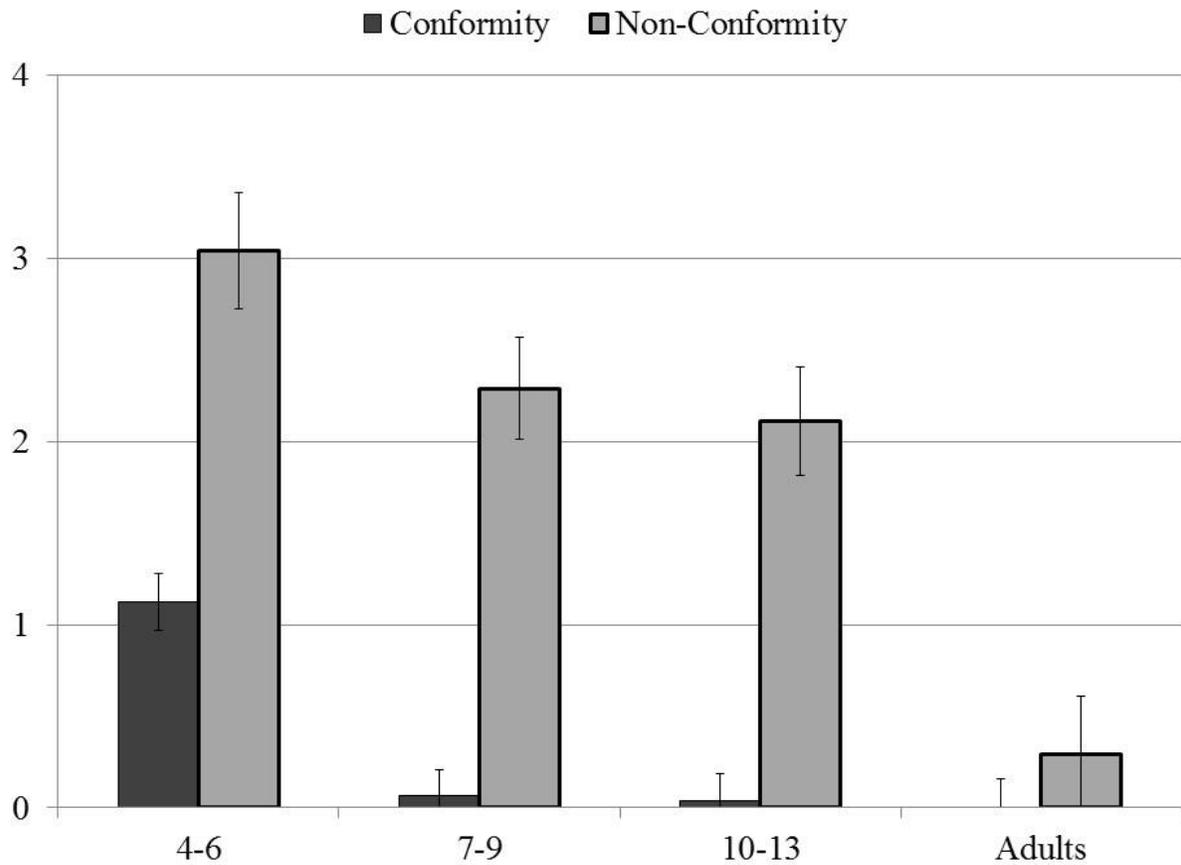


Figure 2. Disapproval data (Chapter II, Study 1). Mean frequency of disapproval of conformity and non-conformity across age groups. Scores could range from 0 to 4. Error bars depict standard errors.

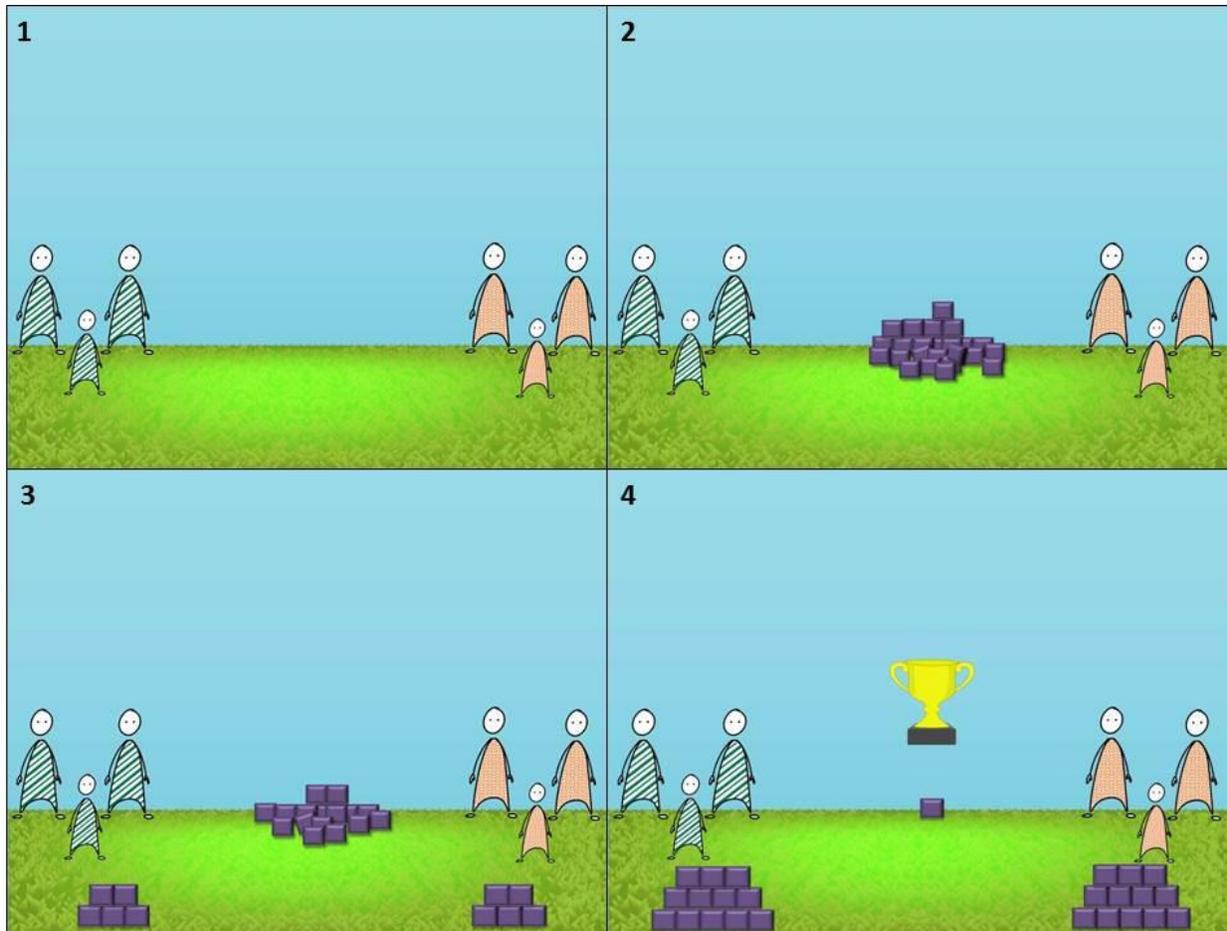


Figure 3. Competition vignette (Chapter II, Study 2).

Slide 1: “I’m going to tell you about two kinds of groups. This group is called Hibbles (pointing to left), and this group is called Glerks (pointing to right). Can you point to the Hibbles? Can you point to the Glerks?”

Slide 2: “The Hibbles and the Glerks had a big contest against each other. They had a contest to see which group could build the tallest tower with these blocks. The Hibbles worked really hard to build a tall tower really fast and the Glerks worked really hard to build a tall tower really fast too! They each ran to get more and more blocks for their towers. Their towers got taller...”

Slide 3: “...and taller.”

Slide 4: “Each group worked against each other so they could make their tower really tall. Whichever group builds the tallest tower is going to get a really great prize. There weren’t enough blocks for everyone, so only one group would get it in the end.”

Comprehension Check: “Were there enough blocks for everyone?” [if answered correctly; “Right, there weren’t enough blocks for everyone, so only one group would win and get a prize in the end.”]

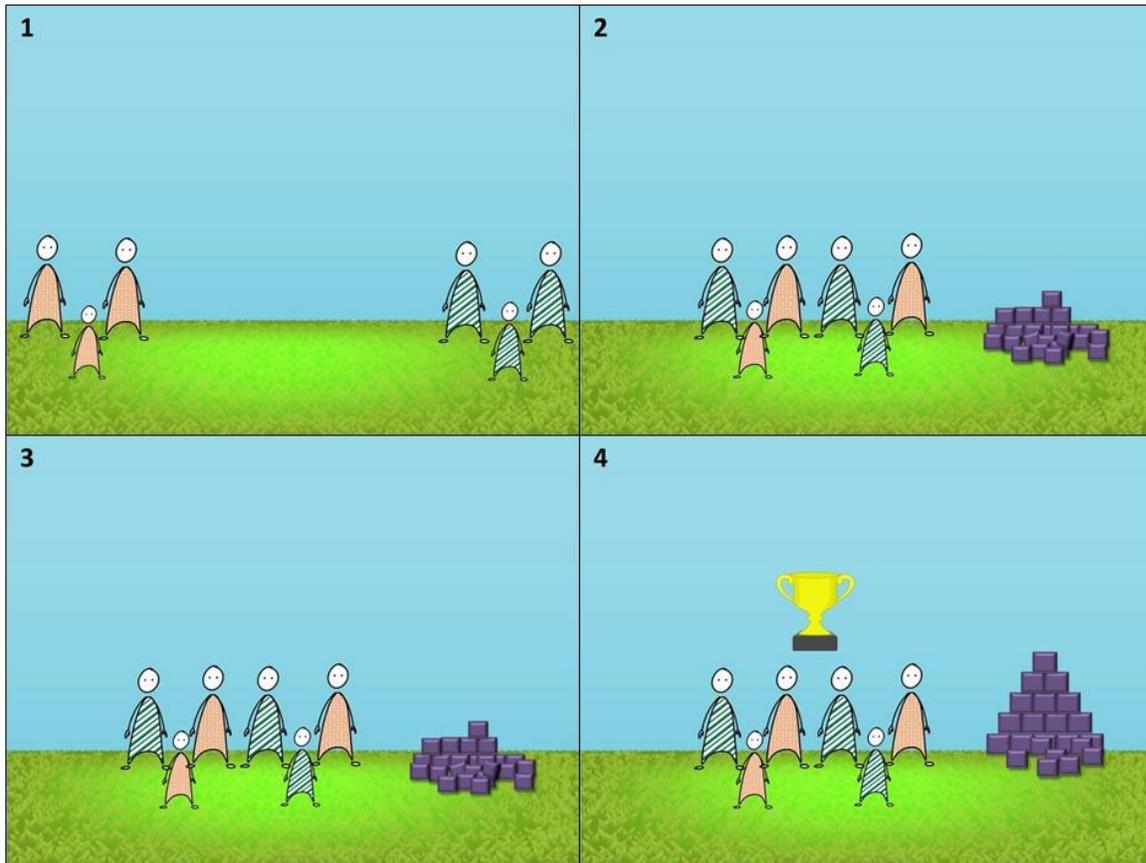


Figure 4. Cooperation vignette (Chapter II, Study 2).

Slide 1: “I’m going to tell you about two kinds of groups. This group is called Hibbles (pointing to left), and this group is called Glerks (pointing to right). Can you point to the Hibbles? Can you point to the Glerks?”

Slide 2: “The Hibbles and the Glerks played a big game together. They played a game where they built a really tall tower with these blocks. The Hibbles worked really hard to build a tall tower really fast and the Glerks worked really hard to build a tall tower really fast too! They each ran to get more and more blocks for the tower. The tower got taller...”

Slide 3: “...and taller.”

Slide 4: “Each group worked together so they could make the tower really tall. When the groups finish the tower together, they are both going to get a really great prize. There were enough blocks for everyone, so both groups would get a prize in the end.”

Comprehension Check: “Were there enough blocks for everyone?” [if answered correctly; “Right, there were enough blocks for everyone, so both groups would win and both groups would get a prize in the end.”

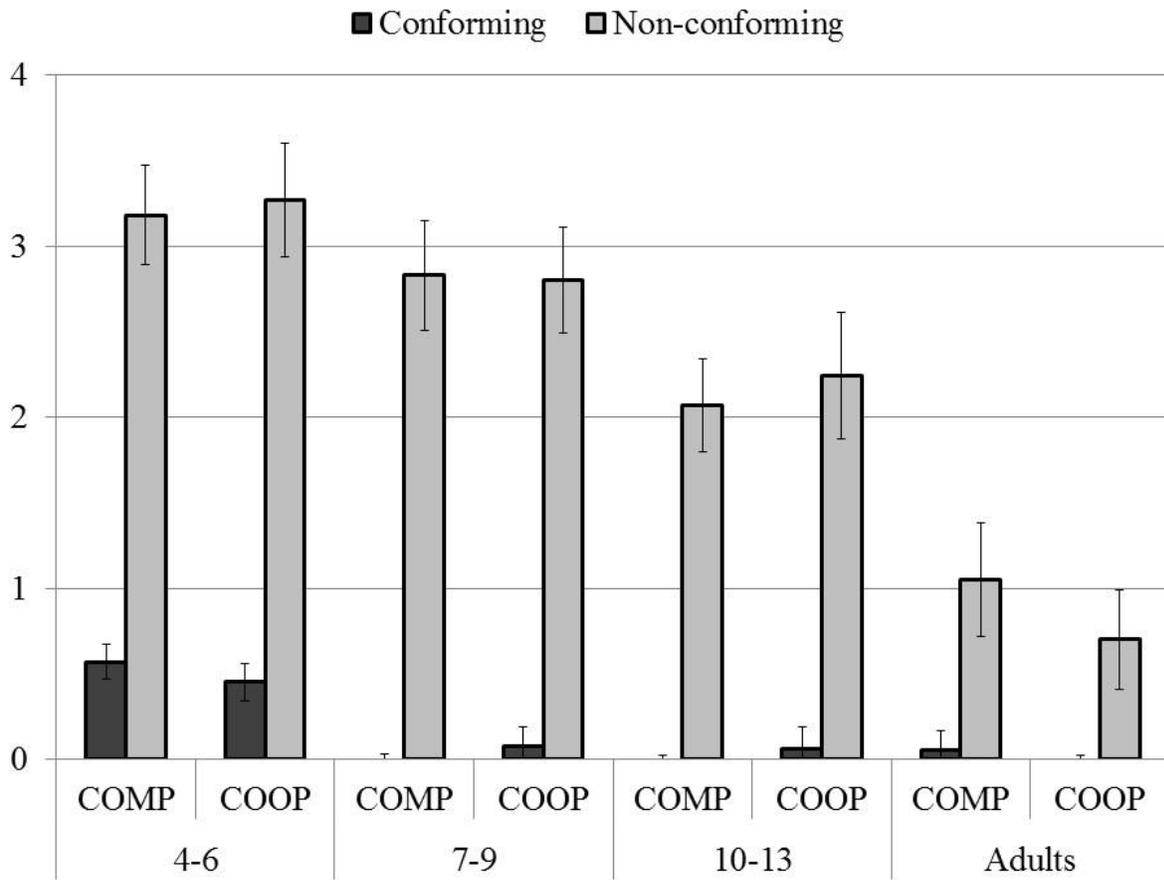


Figure 5. Disapproval data (Chapter II, Study 2). Mean frequency of disapproval of conformity and non-conformity across age groups and conditions. Scores could range from 0 to 4. Error bars display standard errors.

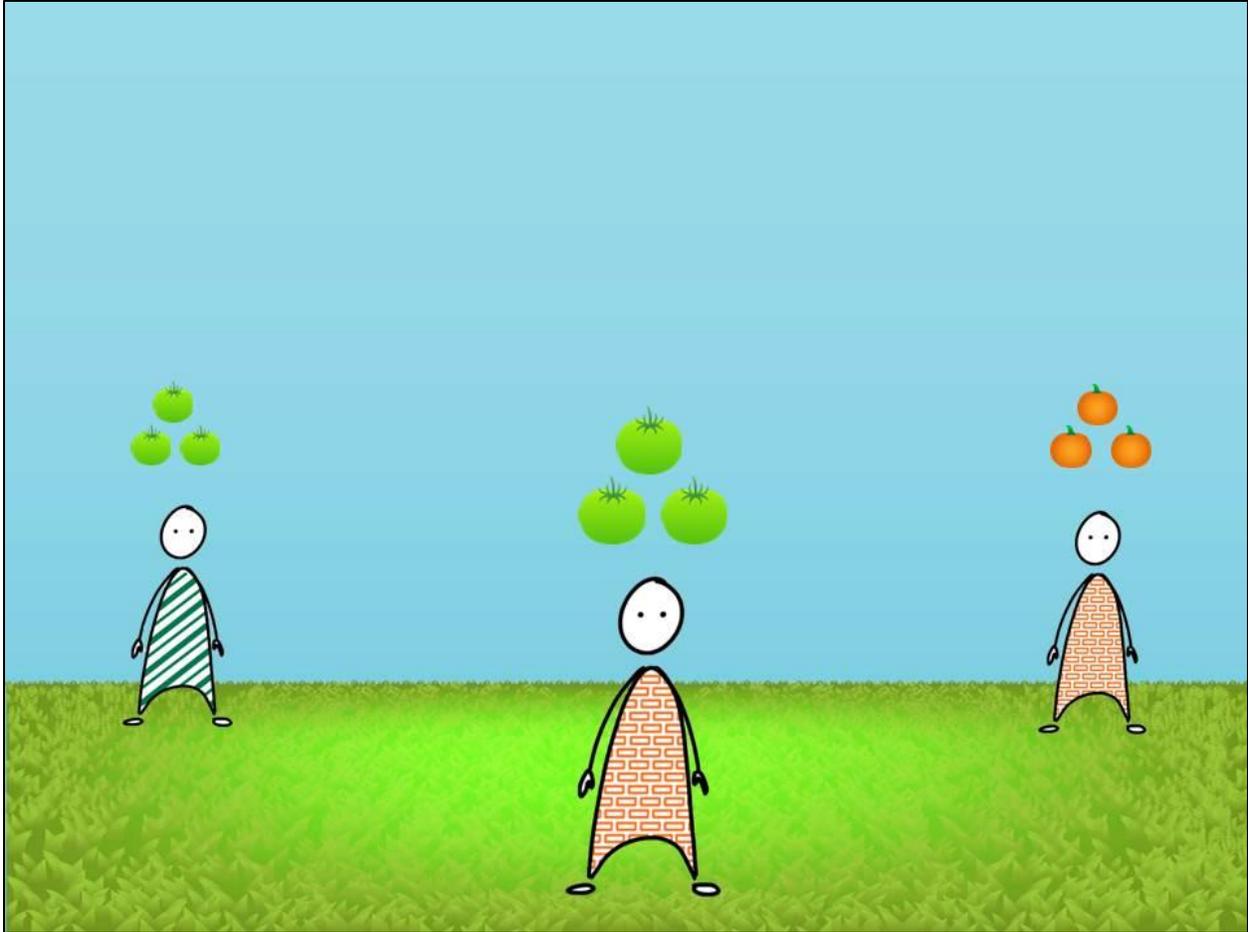


Figure 6. Example trial (Chapter II, Study 3).

“This one (pointing to left) eats these kind of berries and this one (pointing to right) eats these kind of berries. Look, this one (pointing to target) eats these kind of berries.

Evaluation: “Is it okay or not okay for this one to eat these kind of berries?”

Negativity: [if participant evaluated this as not okay] “Is it a little bad, pretty bad, or very, very bad, for this one to eat these kind of berries?”

Explanation: “Why is it [okay or not okay] for this one to eat these kind of berries?”

CHAPTER III

Minimal perceptual and linguistic cues foster children's prescriptive judgments

Abstract

In Chapter 2, I found that children used descriptive regularities of social groups (what *is*) to generate prescriptive judgments (what *should be*). In this chapter, I examined whether this tendency held when the regularities were introduced through group presence, category labels, or generic statements. Children (4-9 years; $N = 203$) were randomly assigned to one of four conditions that manipulated how descriptive group regularities were presented: group presence (e.g., “These ones [a group of three individuals] eat this kind of berry”), category labels (e.g., [showing an individual] “This Hibble eats this kind of berry”), generic statements (e.g., [showing an individual] “Hibbles eat this kind of berry”), or control (e.g., [showing an individual] “This one eats this kind of berry”). Then, children saw conforming and non-conforming individuals and were asked to evaluate their behavior. As predicted, children evaluated non-conformity negatively in all but the control condition, demonstrating that minimal perceptual and linguistic cues provoked children to treat social groups as having normative force.

Introduction

As mentioned in Chapter 2, an important open question was what information signaled to children that a feature was group-relevant, and thus normative. Roberts et al. (2017a) provided children with several converging cues: the groups consisted of individuals in sets of three (thereby highlighting group presence), they were labeled with a common noun (i.e., Hibble, Glerk), and they were described with generic statements in which properties were attributed to

categories (e.g., “Hibbles eat these kinds of berries”). When these cues were present, children made prescriptive judgments; when they were absent, children did not. Thus, although children used group regularities to generate prescriptive judgments, it remains unclear which cues or combination of cues shift a behavior from being a descriptive “is” to a prescriptive “should.” That is, are prescriptive judgments elicited by group presence, category labels, generic statements, or some combination of these three factors?

Seeing a group of individuals engage in a common behavior has profound effects on our social cognition. Take as an example Asch's (1955) classic research on social pressure (i.e., normative social influence): Individuals confronted with three or more people who share a belief, feel the pressure to also hold that belief, even if they suspect the belief to be false. Thus, because the presence of a group is a strong predictor of conformity, one possibility is that simply seeing a group of individuals that share a common behavior will generate prescriptiveness (i.e., all individual group members *should* share that behavior). Indeed, recent research suggests that being exposed to a group of individuals who share common properties increases the likelihood of that group being perceived as coalitional, as well as the likelihood of stereotyping individuals within that group (e.g., Bigler, Jones, & Lobliner, 1997; Cosmides, Tooby, & Kurzban, 2003).

Group labels (i.e., count nouns; Macnamara, 1982), even when only an individual group member is present, play a critical role in shaping young children's categorization. Labels make categories more salient, encourage children to form new categories and treat them as stable, and promote category-based inferences (Baron, Dunham, Banaji, & Carey, 2014; Gelman & Markman, 1986; Graham, Keates, Vukatana, & Khu, 2013; Waxman & Markow, 1995). For example, Waxman (2010) introduced children to individuals (e.g., Black woman, White man), provided novel properties about those individuals (e.g., “is good at a game called zaggit”), and

assessed whether participants expected newly encountered individuals to share the novel property. When individuals received category labels (e.g., “This one is a *Wayshan*”), children expected newly encountered same-race and same-gender individuals to share the same property, whereas when individuals were not labeled (e.g., “This one *eats big lunches*”), they did not expect newly encountered same-race or same-gender individuals to share the same property. Thus, labels facilitated group-based inferences. However, this finding does not speak to whether labels, in addition, generate children’s prescriptive judgments toward novel groups.

Labels may license stronger inferences when they are coupled with generic statements (e.g., expressions that refer to categories). For example, the generic statement “Hibbles eat this kind of berry” may license stronger inferences than the specific statement “These Hibbles eat this kind of berry”, because the generic implies that the category is closely linked to the property, general across time and contexts, and indicative of commonalities shared among individual group members (Cimpian & Erickson, 2012; Gelman et al., 2010; Graham, Nayer, & Gelman, 2011; Prasada, 2000). Indeed, when 4-year-olds were presented with a completely novel category (i.e., Zarpies), they were more likely to make category-based inferences when Zarpies were introduced via generics (e.g., “Zarpies like to sing”) than when they were introduced via specific labels (e.g., “This Zarpie likes to sing”) or no labels (e.g., “This likes to sing”; Gelman et al., 2010; Rhodes, Leslie, & Tworek, 2012), demonstrating that generic statements, compared to labels alone, generate more category-based inferences, and suggesting they may likewise generate more prescriptive judgments.

The present research tested the extent to which group presence, category labels, and generic statements bolstered children’s prescriptive judgments of group regularities. To do this, we presented children with two kinds of groups, and randomly assigned children to one of four

conditions in which we manipulated how the group regularities were conveyed: (a) group presence (each group included three individuals who were described without labels or generic statements), (b) labels (each group included one individual who was described with a category label but not a generic statement), (c) generic (each group included one individual who was described with a label and a generic statement), or (d) control (each group included one individual who was described without a label or generic statement). Because we wanted group membership to be apparent in all conditions, members of different groups were spatially segregated and wore distinct clothing patterns. Prior research indicates that such cues by themselves do not foster prescriptive judgments (Roberts et al., 2017a, Study 2). We predicted that (1) children would use group presence, category labels, and generic statements to make prescriptive judgments, (2) children would make fewer prescriptive judgments when not provided with these cues (thereby replicating Roberts et al., 2017a), and (3) children would make more prescriptive judgments when hearing generic statements than when hearing only labels. No additional a priori predictions were made regarding the relative effects of group presence, labels, and generic statements. We focused on the age groups of 4-6 and 7-9 years, given that Roberts et al. (2017a) found that this was when important developmental changes occurred on this task, and which other research shows captures significant changes in children's social category concepts (Quintana, 1998; Rhodes & Gelman, 2009; Roberts & Gelman, 2015, 2017a).

Method

Participants

Two age groups of children were included ($N = 203$): 4- to 6-year-olds ($n = 106$, 51% female, M age = 5 years and 6 months, range = 4 years and 1 month - 6 year and 11 months), and 7- to 9-year-olds ($n = 97$; 61% female, M age = 8 years and 4 months, range = 7 years and 0

months to 9 years and 11 months). All children were recruited in the Midwest at two university-affiliated museums. The sample was mostly white/European American (60% white/European American, 12% Asian American/Asian/Pacific Islander, 10% multiracial, 5% black/African American, 3% Hispanic/Latino, and 12% other or not reported).

Materials and Procedure

The task was adapted from Roberts et al. (2017a) and presented via Qualtrics. Children were randomly assigned to one of four conditions: Group presence (i.e., sets of three individuals per group, which were spatially segregated, distinguished by clothing patterns, and presented without category labels or generic statements); Label (one individual per group, spatially segregated, distinguished by clothing patterns, and presented with category labels); Generic (one individual per group, spatially segregated, distinguished by clothing patterns, and presented with category labels and generic statements); and Control (one individual per group, spatially segregated, distinguished by clothing patterns, and presented without labels or generic statements). For sample trials see Figures 7 and 8.

Conditions

Presence. Children (4-6: $n = 25$, 7-9: $n = 24$) were introduced to two novel groups that were presented without category labels or generic statements: “I’m going to tell you about these ones (*pointing*) and these ones (*pointing*).” Each group consisted of three individuals (one set of three located on the left side of the screen, another set located on the right side of the screen), with group membership portrayed by clothing pattern (i.e., orange rectangles vs. green stripes). Next, children received eight test trials. Across all conditions, there were four behavioral domains: Food, Games, Language, and Music, and each behavior matched its corresponding group in color (e.g., green clothing pattern corresponded with a green musical note). On each

trial, children were given regularities for both groups and then shown a conforming or a non-conforming individual. For example, “These ones (*pointing to the group with orange rectangles*) eat this kind of berry and these ones (*pointing to the group with green stripes*) eat this kind of berry. Look (*revealing and pointing to an individual who either conformed or did not conform to their group*), this one is eating this kind of berry.”

Label. Children (4-6: $n = 31$, 7-9: $n = 24$) were introduced to two individuals who were spatially segregated, distinguished by clothing pattern (identical clothing to those in the group presence condition), and presented with contrasting category labels: “I’m going to tell you about these two. This one is a Hibble (*pointing*) and this one is a Glerk (*pointing*).” Next, children were told a fact about each individual and then introduced to other individuals who either conformed or did not conform to the property exhibited by the individual with the same name (eight trials). For example, “This Hibble eats this kind of berry (*pointing*) and this Glerk eats this kind of berry (*pointing*). Look (*revealing and pointing to an individual who either conformed or did not conform to their labeled group*), this Hibble is eating this kind of berry (*pointing*).”

Generic. As in the Label condition, children (4-6: $n = 26$, 7-9: $n = 25$) were introduced to two individuals who were spatially segregated and distinguished by clothing pattern: “I’m going to tell you about these two. This one is a Hibble (*pointing*) and this one is a Glerk (*pointing*).” Next, children were told a fact about each individual and then introduced to conforming or non-conforming individuals (eight trials). Unlike the Label condition, however, the initial facts were provided in the form of generic statements: For example, “Hibbles eat this kind of berry (*pointing*) and Glerks eat this kind of berry (*pointing*). Look (*revealing and pointing to an individual who either conformed or did not conform to their generically-referenced group*), this Hibble is eating this kind of berry (*pointing*).”

Control. Children (4-6: $n = 24$, 7-9: $n = 24$) were introduced to two individuals who were spatially segregated, distinguished by clothing pattern, but presented without category labels or generic statements (see also Roberts et al., 2017a, Study 2): “I’m going to tell you about these two - this one (*pointing*) and this one (*pointing*).” Children were then told a fact about each individual and introduced to conforming and non-conforming individuals. For example, “This one eats to this kind of berry (*pointing*) and this one eats this kind of berry (*pointing*). Look (*revealing and pointing to an individual who either conformed or did not conform to the other individual with the same appearance*), this one is eating this kind of berry (*pointing*).”

Across all conditions, each participant saw eight trials. Four trials depicted a conforming individual and four trials depicted a non-conforming individual. Trials were presented in random order, and across participants the left-right position of the groups/individuals was counterbalanced. In the Label and Generic conditions, we also counterbalanced which label was associated with which clothing pattern.

Measures and coding

First, children were asked whether or not the behaviors of the conforming/non-conforming individuals were “okay” or “not okay” (evaluation, e.g., “Is it okay or not okay for this Hibble to eat this kind of berry?”). We calculated the frequency with which these evaluations occurred for both conformity and non-conformity trials (scores for each could range from 0 to 4), focusing on the frequency of “not okay” evaluations, which reflected disapproval, as the dependent variable. (Note that the frequency of “okay” and “not okay” responses were precise inverses.) Second, children were asked how bad or how good a specific behavior was. Children who evaluated behaviors as “not okay” were presented with a scale of three increasingly unhappy faces and were asked, “Is it a little bad, pretty bad, or very, very bad?” (1 = *a little bad*,

2 = *pretty bad*, 3 = *very, very bad*), and children who evaluated behaviors as “okay” were presented with a scale of three increasingly happy faces and were asked, “Is it a little good, pretty good, or very, very good?” (1 = *a little good*, 2 = *pretty good*, 3 = *very, very good*).

Third, children were asked to explain their evaluations (e.g., “Why is it not okay for this one to eat this kind of berry?”). Responses were coded into five types based on previous research (Roberts et al., 2017a; Rhodes, 2014): (a) norm-based (b) group-based (c) individual-based, (d) similarity-based, and (e) other (see Table 1 for a description of the coding scheme). Codes were not mutually exclusive, so children could appeal to multiple explanation types within a single response. These responses were coded by two research assistants who were blind to hypotheses of the study (Cohen’s kappa = .77) and disagreements were resolved by discussion. We calculated the percentage of times that each type of explanation was provided, out of the total number of trials, for each response type (e.g., a child who disapproved of non-conformity on 4 trials and gave a norm-based explanation on 1 of those trials was coded as having given norm-based explanations for disapproved non-conformity 25% of the time).

As a comprehension check, at the end of the task, children were asked, “What does it mean for something to be not okay?” (open-ended) and “Does 'not okay' mean that someone should or should not do something?” (0 = should, 1 = should not). Only children who indicated norm-based reasoning in their open-ended question (e.g., “you aren’t supposed/allowed to do it”, “it means something bad will happen”) and said that “not okay” means someone “should not” do something were included in the final sample. Seven children did not meet these criteria and were therefore excluded from the final sample.

Parent survey. For exploratory purposes, parents were also given the option of completing a survey (81% response rate) adapted from Feldman (2003) that assessed their views

on authoritarian parenting ($\alpha = .51$), conformity ($\alpha = .41$), and respect for common norms ($\alpha = .61$). However, these measure yielded low reliabilities and were not related with children's responses, and are therefore not reported further.

Results

There were no significant effects of any of the counterbalancing factors or of the behavioral domains, so the data were collapsed across these variables. All significant effects at the $p < .05$ level were followed by Bonferroni corrected pairwise comparisons.

Disapproval toward non-conformity and conformity

We first tested whether children were more disapproving of non-conformity than of conformity, and whether their rates of disapproval varied across conditions. To test this, we conducted a 2 (age group: 4-6, 7-9) x 2 (behavior: conformity, non-conformity) x 4 (condition: Group Presence, Label, Generic, Control) repeated measures ANOVA with behavior as a within-subjects variable, age group and condition as between-subjects variables, and the frequency of not-okay evaluations as the dependent variable. A main effect of behavior, $F(1, 195) = 139.83, p < .001, \eta_p^2 = .42$, showed that non-conformity was indeed disapproved of more frequently than conformity, and a main effect of age group, $F(1, 195) = 10.23, p = .002, \eta_p^2 = .05$, showed that 4- to 6-year-olds were more disapproving than 7- to 9-year-olds. There was also a main effect of condition, $F(3, 195) = 4.93, p = .003, \eta_p^2 = .07$. Planned comparisons showed that compared to the Control condition, disapproval was higher in the Generic condition ($p = .001$). There was also a significant interaction of behavior and condition, $F(3, 195) = 8.34, p < .001, \eta_p^2 = .11$. Planned comparisons revealed that non-conformity was disapproved more frequently than conformity in the Group Presence, Label, and Generic conditions ($ps < .001$), but not in the Control condition. For conformity, disapproval did not differ significantly between any of the conditions ($ps =$

1.00). For non-conformity, disapproval was lowest in the Control condition compared to all other three conditions ($ps \leq .012$), and disapproval did not differ across the three experimental conditions ($ps > .55$). There were no interactions involving age group.

We next conducted one-sample t -tests to test responses against chance (i.e., 2). When looking at disapproval for conformity, both age groups across all four conditions were below chance (thereby indicating approval). When looking at disapproval for non-conformity, 4- to 6-year-olds in the Generic condition were above chance, 4- to 6-year-olds in the Control condition were marginally below chance, and 7- to 9-year-olds in the Control condition were below chance. All other groups were at chance. These statistics are presented in Table 7 and the data are presented graphically in Figure 9. Non-parametric Wilcoxon signed-ranks tests provided further insight into the chance-level responses (see Table 8 for all data and statistics). Overall, these analyses revealed that in the experimental conditions, regarding non-conformity, half of the children most often approved and half of the children most often disapproved (though this was not the case for 4- to 6-year-olds in the Generic Statement condition). See the General Discussion for insight into these response patterns.

Negativity toward non-conformity

Next, we focused only on the children who disapproved of non-conformity on at least one trial and were asked how bad the behavior was (1 = *kind of bad*, 2 = *pretty bad*, 3 = *very, very bad*; $n = 138$). A univariate ANOVA with age group (2: 4-6, 7-9) as a between-subjects variable and negativity as the dependent variable (i.e., average rating score across non-conformity trials on which children indicated “not okay”; scores could range from 1 to 3) yielded a main effect of age group, $F(1, 130) = 21.46, p < .001, \eta_p^2 = .14$, showing that 4- to 6-year-olds ($M = 2.47, SE = .08$) were more negative than 7- to 9-year-olds ($M = 1.85, SE = .09$). This effect was qualified by

a significant interaction of age group and condition, $F(3, 138) = 3.76, p = .02, \eta_p^2 = .07$. Four-to six-year-olds were more negative than 7- to 9-year-olds in the Control condition (4-6: $M = 2.70, SE = .18$; 7-9: $M = 1.34, SE = .22, p < .001$) and Group Presence condition (4-6: $M = 2.66, SE = .15$; 7-9: $M = 2.09, SE = .17, p = .012$). The two age groups did not differ in their negativity in either the Label condition (4-6: $M = 2.19, SE = .14$; 7-9: $M = 2.00, SE = .17, p = .41$) or the Generic condition (4-6: $M = 2.33, SE = .14$; 7-9: $M = 1.93, SE = .15, p = .057$), with ratings in both age groups corresponding roughly to the evaluation that the non-conformity was “pretty bad”.

Positivity toward non-conformity

We next focused only on the children who approved of non-conformity on at least one trial and were asked how good the behavior was (1 = *kind of good*, 2 = *pretty good*, 3 = *very, very good*; $n = 125$). A univariate ANOVA with age group (2: 4-6, 7-9) as a between-subjects variable and positivity as the dependent variable (i.e., average rating score across non-conformity trials on which they indicated “okay”; scores could range from 1 to 3) yielded only a main effect of age group, $F(1, 117) = 4.37, p = .039, \eta_p^2 = .04$, showing that 4- to 6-year-olds were more positive than 7- to 9-year-olds (4-6: $M = 2.33, SE = .07$; 7-9: $M = 2.09, SE = .08$) when approving of non-conformity.

Positivity toward conformity

Lastly, we focused on the children who approved of conformity on at least one trial and were asked how good the behavior was (1 = *kind of good*, 2 = *pretty good*, 3 = *very, very good*; $n = 196$). A univariate ANOVA with age group (2: 4-6, 7-9) as a between-subjects variable and positivity as the dependent variable (i.e., average rating score across conformity trials on which they indicated “okay”; scores could range from 1 to 3) yielded a main effect of age group, $F(1,$

188) = 8.71, $p = .004$, $\eta_p^2 = .04$, showing that 4- to 6-year-olds were more positive than 7- to 9-year-olds (4-6: $M = 2.59$, $SE = .05$; 7-9: $M = 2.38$, $SE = .05$) when approving of conformity.

Explanations

We next analyzed the explanations that children provided after they were asked why they approved or disapproved of a given behavior (e.g., “Why is it okay for this one to eat this kind of berry?”). Explanations were given for three response types: Disapproved non-conformity, approved non-conformity, and approved conformity. Responses for disapproved conformity were not analyzed because this response was rarely given. We focused on the frequency of the four primary explanation types (norm-based, group-based, individual-based, similarity-based) within each response type (e.g., approved conformity), not including “other” miscellaneous explanations, and did not statistically compare across response type because not all participants provided each type of response. The percentage of times each explanation type was provided, out of the total number of trials, was analyzed via repeated measures ANOVAs in which age group and condition were between-subjects variables, explanation type was a within-subjects variable, and the percentage of given explanations was the dependent variable. The Huynh-Feldt correction, which adjusts the degrees of freedom for the calculated F values (Field, 2011; Huynh & Feldt, 1976), was used because the explanation data violated the repeated measures assumption of sphericity. See Tables 9 and 10 for all explanation data.

Explanations about disapproved non-conformity

Focusing on children who evaluated non-conformity as not okay ($n = 130$), we found a significant main effect of age group, $F(1, 122) = 6.70$, $p = .011$, $\eta_p^2 = .05$, a significant difference in explanation type, $F(2.77, 337.62) = 17.05$, $p < .001$, $\eta_p^2 = .12$, an interaction of age group and explanation type, $F(2.67, 337.62) = 4.20$, $p = .006$, $\eta_p^2 = .03$, an interaction of explanation type

and condition, $F(8.30, 337.62) = 3.55, p < .001, \eta_p^2 = .08$, and an interaction of age group, explanation type, and condition, $F(8.30, 337.62) = 1.96, p = .048, \eta_p^2 = .05$. Four- to six-year-olds gave mostly similarity-based explanations, and they did so at similar rates across all conditions, whereas 7- to 9-year-olds gave mostly group-based explanations and they did so most often in the Label and Generic conditions (though in the Group Presence condition they gave mostly similarity-based explanations).

Explanations about approved non-conformity

Focusing on participants who approved of non-conformity ($n = 118$), we found a significant main effect of age group, $F(1, 110) = 9.71, p = .002, \eta_p^2 = .08$, a significant difference in explanation type, $F(2.16, 237.53) = 26.40, p < .001, \eta_p^2 = .19$, and a significant interaction of age group and explanation type, $F(2.16, 237.53) = 2.16, p = .001, \eta_p^2 = .07$. Both age groups gave mostly individual-based explanations, though 7- to nine-year-olds did so to a greater extent than did 4- to 6-year-olds.

Explanations about approved conformity

Focusing on participants who approved of conformity ($n = 195$), we found an interaction of age group and explanation type, $F(2.86, 535.38) = 3.27, p = .021, \eta_p^2 = .02$. The 4- to 6-year-old group gave more similarity-based explanations than norm-based explanations, and the 7- to 9-year-old group gave more group-based, individual-based, and similarity-based explanations than norm-based explanations.

Discussion

Children take what *is* to infer what *should be* (Kenward, 2012; Rakoczy & Schmidt, 2013), even with regard to unfamiliar, third-party groups engaged in morally neutral behaviors (Roberts et al., 2017a). That is, if young children are shown a group that is described as sharing a

common property (i.e., descriptive regularity), children infer that individual group members should share that property and that it is bad if they do not (i.e., prescriptive judgment). In the present research, we tested the extent to which this descriptive-to-prescriptive tendency was facilitated by attention to group presence, category labels, or generic statements. Indeed, when group regularities were conveyed by showing multiple individuals engaged in a common behavior (i.e., group presence), words that highlighted an individuals' group membership (i.e., category labels), or statements that associated group membership with specific properties (i.e., generic statements), children disapproved of non-conformity (consistent with our first prediction). In a control condition that included none of these factors, however, children were significantly less likely to disapprove of non-conformity (consistent with our second prediction), demonstrating that when the emphasis was on individuals and not groups, children were less prescriptive (see also Study 2 in Roberts et al., 2017a). The ease with which group-based, prescriptive reasoning is elicited, combined with its early emergence in younger children, suggests that such reasoning may be a fundamental aspect of human social cognition.

We also note that the generic statement condition was the only condition in which 4- to 6-year-olds disapproved of non-conformity above chance levels, and in which they disapproved more often than they approved. At the same time, response patterns in the generic condition did not differ significantly from those in the other experimental conditions. Future research would be needed to determine whether or not generic statements may be more powerful than group presence or category labels alone in fostering prescriptive judgments (consistent with our third prediction and with previous research; Gelman et al., 2010; Rhodes et al., 2012). For example, a more fine-grained scale to assess the degree of negativity that children feel toward non-conformity may be more sensitive to condition differences in this age group. At minimum, these

data are consistent with prior research indicating that generics help children learn about groups, guide their expectations about individuals, and help them generate inferences about future behaviors (e.g., Cimpian & Erickson, 2012; Gelman, 2003, Gelman et al., 2010; Graham et al., 2011; Prasada, 2000), and further suggest that generics may preferentially license children's prescriptive judgments.

One unexpected finding was that younger children rarely used norm-based reasoning (e.g., "They shouldn't do that") to explain why they disapproved of non-conformity, in contrast to earlier work (Roberts et al., 2017a). One possibility is that, although group presence, category labels, and generic statements elicit disapproval and negativity, they may only be collectively strong enough to elicit norm-based justifications; perhaps a combination of the three conveys a stronger regularity that licenses children to make strong normative justifications. Future work is needed to test this empirically. For now, though, the present data indicate that group presence, category labels, and generic statements license prescriptiveness in the form of disapproval and negativity.

Also of note, younger children tended to respond in a more extreme manner than older children: they were more disapproving than older children, on both conformity and non-conformity trials, and were more negative on the negativity scale and more positive on the positivity scale. One interpretation of these patterns is that younger children are more attentive to both conformity and non-conformity, and therefore feel especially positive about conformity and especially negative about non-conformity. This interpretation aligns with recent research suggesting that 3-year-olds are significantly more troubled by norm violations than are 5-year-olds (Hardecker et al., 2016). Regardless, of primary interest here was the directionality of children's responses, which varied as a function of the specific behaviors they saw (i.e., non-

conformity vs. conformity). That is, younger children's more extreme responses are independent of the finding that they were systematic in their responses: in the experimental conditions, they were more disapproving of non-conformity than of conformity, and more approving of conformity than of non-conformity.

An important question for future research is why the children interpret group presence, labels, and generic statements as indicative of what is normatively correct. One possibility is that each of these features provides children with the information that a given property is systematically shared among a group of individuals, rather than possessed by only one. Once children take note of a descriptive regularity, they may reason that adherence and conformity to group regularities is critical for individual and group functioning (Tomasello, 2016), and subsequently interpret the regularity as having prescriptive force. In other words, children may be prepared to detect group regularities, which could be signaled multiply by group presence, category labels, and generic statements.

A second possibility (not mutually exclusive from the first) is that such cues encourage children to essentialize the social group in question (see Bigler & Liben, 2006 for a related theoretical model). For example, when children see a group of individuals who share common features (e.g., people with dark skin live on the other side of town), they may come to conceptualize those features as essential to their identity, and subsequently use group membership as a basis for prescriptiveness (e.g., people with dark skin *should* live on the other side of town). The explanation data support this possibility, showing that in the absence of labels, children justified their prescriptiveness through similarity-based explanations. Similarly, labels and generic statements could signal to children that that group membership is salient and important, subsequently influencing them to interpret group membership as essential and

therefore prescriptive. Indeed, in both of these conditions, both age groups most often appealed to group-based explanations (i.e., they often appealed to the category label, which was given to them in both of these conditions). This finding is consistent with recent research suggesting that when given race-based labels, children (ages 5 to 10) are especially likely to essentialize race (Roberts & Gelman, 2017b). Future research is needed to systematically test the extent to which essentialist reasoning mediates the descriptive-to-prescriptive tendency detected here, which may also help understand the non-random individual response patterns. Recall from the non-parametric tests that children were often split in their evaluations of non-conformity: half of the children most often disapproved, whereas the other half most often approved (excluding 4- to 6-year-olds in the generic statement condition). One possibility is that children who interpreted the properties as “essential” to the group were more likely to disapprove of non-conformity. Indeed, recent work suggests that children who attribute behaviors to groups (e.g., Hibbles eat that kind of berry because of something about *Hibbles*), rather to properties (e.g., Hibbles eat that kind of berry because of something about that *berry*), are more likely to conceptualize non-conformity as a norm violation (see Tworek & Cimpian, 2016).

In conclusion, the data presented in Chapter 3 contribute to a growing body of literature on children’s reasoning about norms (Engelmann et al., 2016; Riggs & Young, 2016; Schmidt, Hardecker, et al., 2016), showing that such reasoning influences children’s descriptive-to-prescriptive tendency even under minimal contexts with minimal cues to group regularities. Stated more plainly, when children see a group of individuals, hear a category label, or hear that label explicitly associated with behaviors, they believe that individuals within that group *should* conform, and that it is bad if they don’t. Nevertheless, because the data in Chapter 2 and 3 were derived from exclusively U.S. children (see also Conry-Murray & Turiel, 2012; Cooley &

Killen, 2015; Josephs, Kushnir, Gräfenhain, & Rakoczy, 2016; Kalish, 2012), and therefore limited my understanding of children's normative reasoning across human populations, I conducted a cross-cultural comparison in order to yield more nuanced insights (see Chapter 4).

Table 7.

Disapproval data (Chapter III). Means, standard errors, and one-sample *t*-test statistics comparing average disapproval frequencies against chance (i.e., 2) across age groups, conditions, and behavior type.

Age	Condition	Behavior	M(SE)	<i>t</i>	<i>p</i>	<i>d</i>
4-6	Presence	Non-conformity	2.44(.34)	1.29	.21	.26
		Conformity	0.64(.21)	-6.56	< .001	1.31
	Label	Non-conformity	2.06(.29)	.23	.823	.04
		Conformity	0.81(.21)	-5.56	< .001	.99
	Generic	Non-conformity	2.85(.31)	2.71	.012	.53
		Conformity	0.92(.28)	-3.81	.001	.75
	Control	Non-conformity	1.38(.32)	-1.97	.061	.40
		Conformity	1.04(.24)	-3.92	.001	.80
7-9	Presence	Non-conformity	2.25(.37)	.67	.509	.13
		Conformity	0.17(.10)	-18.65	< .001	3.81
	Label	Non-conformity	2.13(.39)	.32	.755	.06
		Conformity	0.25(.11)	-16.13	< .001	3.29
	Generic	Non-conformity	2.44(.34)	1.27	.217	.25
		Conformity	0.16(.07)	-24.59	< .001	4.92
	Control	Non-conformity	0.75(.27)	-4.62	< .001	.94
		Conformity	0.25(.15)	-11.63	< .001	2.37

Table 8.

Non-parametric data (Chapter III). Number of participants who more often approved (“okay”), more often disapproved (“not okay”), or approved and disapproved equally (Tie), as a function of age group, condition, and behavior.

Age	Condition	Behavior	OK	Not OK	Tie	Z	<i>p</i>	<i>r</i>
4-6	Presence	Non-conformity	10	14	1	1.50	.13	.30
		Conformity	20	1	4	-3.81	<.001	.76
	Label	Non-conformity	12	13	6	.239	.81	.04
		Conformity	24	3	4	-3.79	<.001	.68
	Generic	Non-conformity	7	18	1	2.51	.012	.49
		Conformity	20	5	1	-3.06	.002	.60
	Control	Non-conformity	16	7	1	-1.79	.07	.37
		Conformity	16	3	5	-3.07	.002	.63
7-9	Presence	Non-conformity	9	12	3	.688	.49	.14
		Conformity	23	0	1	-4.63	<.001	.95
	Label	Non-conformity	12	12	0	-.56	.57	.11
		Conformity	23	0	1	-4.51	<.001	.92
	Generic	Non-conformity	9	14	2	1.36	.17	.28
		Conformity	25	0	0	-4.72	<.001	.94
	Control	Non-conformity	21	3	0	-3.12	.002	.64
		Conformity	22	1	1	-4.58	<.001	.93

Table 9.

Explanation data (Chapter III, younger children). Percentage of 4- to 6-year-olds' explanation types for each behavior, across conditions and evaluations types, separately for "not okay" and "okay" responses.

Percentage of Explanation Types								
M(SE)								
Age	Condition	Behavior	Evaluation	N	Norm	Group	Individual	Similarity
4-6	Presence	Non-conformity	Not okay	20	6(7)	25(8)	10(4)	35(9)
		Non-conformity	Okay	11	0	5(5)	19(10)	14(9)
		Conformity	Okay	24	10(5)	17(7)	15(6)	29(7)
	Label	Non-conformity	Not okay	23	18(7)	22(7)	7(3)	31(8)
		Non-conformity	Okay	21	1(3)	4(4)	33(7)	18(7)
		Conformity	Okay	29	7(4)	17(6)	25(6)	31(7)
	Generic	Non-conformity	Not okay	22	18(7)	11(7)	2(3)	33(8)
		Non-conformity	Okay	10	15(5)	9(6)	15(11)	21(10)
		Conformity	Okay	25	8(5)	17(7)	7(6)	25(7)
	Control	Non-conformity	Not okay	11	15(9)	0	2(5)	41(12)
		Non-conformity	Okay	19	1(3)	0	12(8)	20(7)
		Conformity	Okay	23	15(5)	23(7)	8(6)	19(7)

Note. Scores represent percentage of each explanation type out of the total number of trials. 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 (which is why the percentages can add to more than or less than 100). Data for disapproved ("not okay") conformity are not presented because this response was rarely given.

Table 10.

Explanation data (Chapter III, older children). Percentage of 7- to 9-year-olds' explanation types for each behavior, across conditions and evaluation types, separately for "not okay" and "okay" responses.

Percentage of Explanation Types								
M(SE)								
Age	Condition	Behavior	Evaluation	N	Norm	Group	Individual	Similarity
7-9	Presence	Non-conformity	Not okay	16	22(8)	41(9)	2(4)	47(10)
		Non-conformity	Okay	12	2(4)	19(5)	56(10)	25(10)
		Conformity	Okay	22	6(5)	23(7)	31(7)	23(8)
	Label	Non-conformity	Not okay	13	23(9)	69(10)	0	21(11)
		Non-conformity	Okay	13	0	10(5)	47(10)	8(9)
		Conformity	Okay	24	10(5)	38(7)	30(6)	17(7)
	Generic	Non-conformity	Not okay	18	21(7)	49(8)	2(4)	10(09)
		Non-conformity	Okay	11	9(4)	14(5)	30(10)	7(9)
		Conformity	Okay	25	8(5)	21(7)	14(6)	22(7)
	Control	Non-conformity	Not okay	7	7(11)	0	14(6)	54(15)
		Non-conformity	Okay	21	5(3)	0(4)	54(7)	15(7)
		Conformity	Okay	23	0	11(7)	33(6)	22(7)

Note. Scores represent percentage of each explanation type out of the total number of trials. 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 (which is why the percentages can add to more than or less than 100). Data for disapproved ("not okay") conformity are not presented because this response was rarely given.

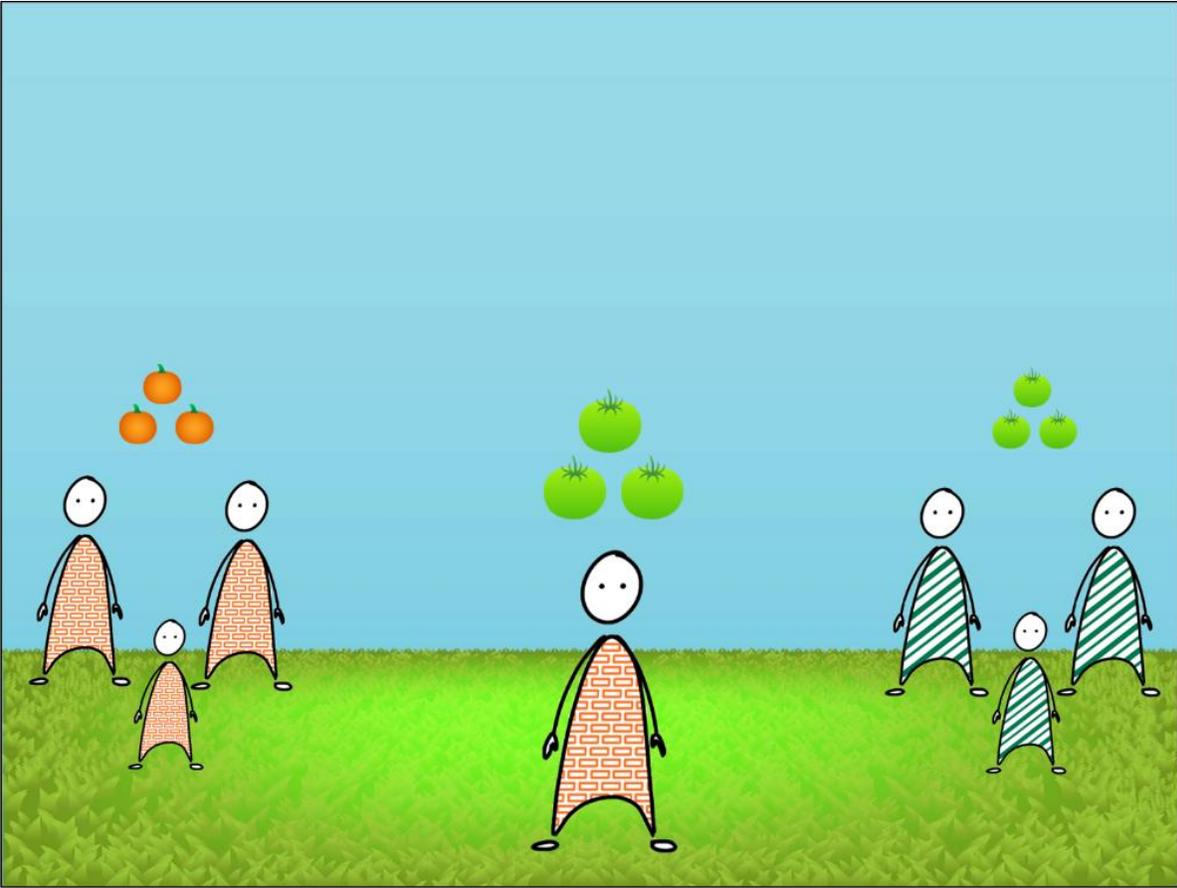


Figure 7. Example trial (Chapter III, Group Presence condition).

“These ones (*pointing to the group with the orange rectangles*) eat this kind of berry and these ones (*pointing to the group with the green stripes*) eat this kind of berry. Look, this one (*pointing to target*) is eating this kind of berries.

Evaluation: “Is it okay or not okay for this one to eat this kind of berry?”

Negativity: [if child evaluated this as not okay] “Is it a little bad, pretty bad, or very, very bad, for this one to eat this kind of berry?”

Explanation: “Why is it [okay or not okay] for this one to eat this kind of berry?”

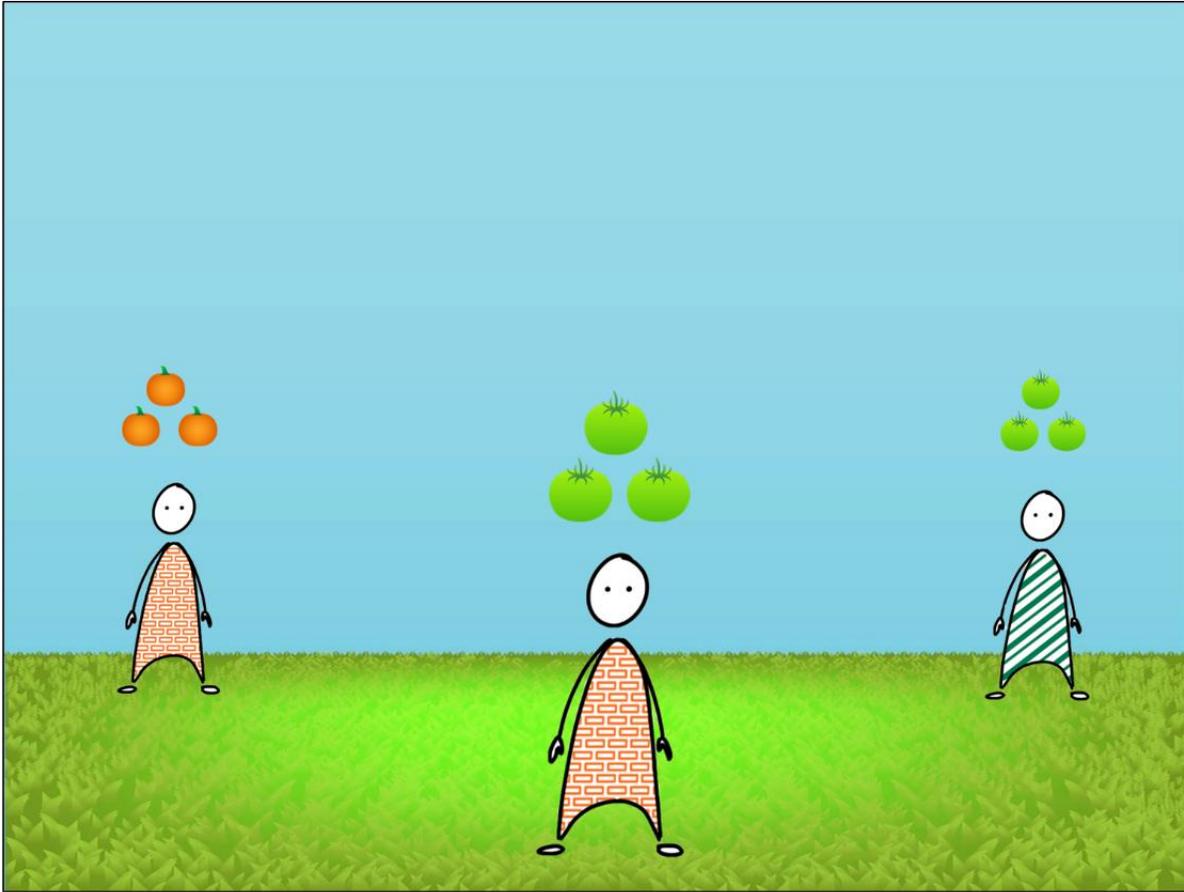


Figure 8. Example trial (Chapter III, Label, Generic, and Control conditions).

Label Condition: “I’m going to tell you about these two. This one is a Hibble (*pointing to the individual with orange rectangles*) and this one is a Glerk (*pointing to the one with green stripes*). This Hibble eats this kind of berry (*pointing*) and this Glerk eats this kind of berry (*pointing*). Look (*revealing and pointing to an individual who either conformed or did not conform to their labeled group*), this Hibble is eating this kind of berry (*pointing*).”

Generic Condition: “I’m going to tell you about these two. This one is a Hibble (*pointing to the individual with orange rectangles*) and this one is a Glerk (*pointing to the individual with green stripes*). Hibbles eat this kind of berry (*pointing*) and Glerks eat this kind of berry (*pointing*). Look (*revealing and pointing to an individual who either conformed or did not conform to their generically-referenced group*), this Hibble is eating this kind of berry (*pointing*).”

Control Condition: “I’m going to tell you about these two. This one (*pointing to the individual with orange rectangles*) and this one (*pointing to the individual with green stripes*). This one eats this kind of berry (*pointing*) and this one eats this kind of berry (*pointing*). Look (*revealing and pointing to an individual either conformed or did not conform to the other individual with the same appearance*), this one is eating this kind of berry (*pointing*).”

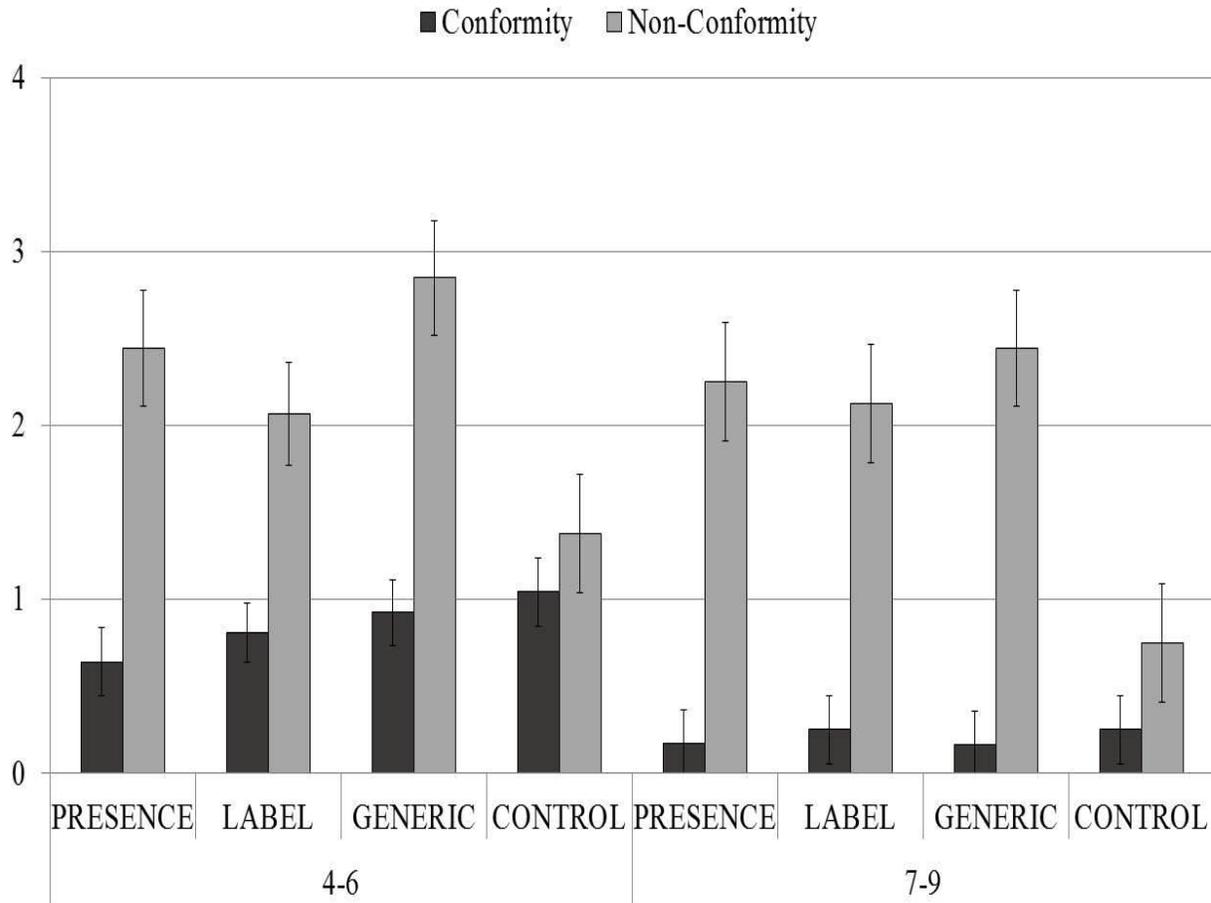


Figure 9. Disapproval data (Chapter III). Mean frequency of disapproval of conformity and non-conformity across age groups and conditions. Scores could range from 0 to 4. Error bars display standard errors.

CHAPTER IV

Children's descriptive-to-prescriptive tendency replicates (and varies) cross-culturally

Abstract

Because the previous studies were conducted with U.S. samples, only, I assessed here whether the descriptive-to-prescriptive tendency held in a sample of children (ages 4 to 13) and adults (ages 18 to 40) from mainland China. Participants were introduced to novel groups (i.e., Hibbles and Glerks) who engaged in contrasting morally neutral behaviors (e.g., listening to different kinds of music) and then to conforming and non-conforming individuals (e.g., a Hibble who listened to music more typical of Glerks). Like U.S. children, Chinese children disapproved of non-conformity and rates of disapproval declined with age. However, compared to U.S. children, younger Chinese children (ages 4 to 6) rated non-conformity more disapprovingly, and unlike U.S. adults, Chinese adults rated non-conformity more negatively than conformity. Moreover, compared to U.S. participants, Chinese participants across all age groups more often appealed to norm-based explanations when justifying their disapproval. These data provide a cross-cultural replication of children's descriptive-to-prescriptive tendency but also reveal cross-cultural variation, and they have implications for understanding the mechanisms that underlie stereotyping and norm enforcement.

Introduction

The previous studies provided new insight into children's normative reasoning and highlighted the profound implications group concepts have for stereotyping and prescriptive

reasoning. If a group is characterized by a property, children believe that individual group members *should be* characterized by that property. The readiness, robustness, and ease with which children's descriptive-to-prescriptive tendency is elicited may be evolutionarily rooted in humans' group-based way of life. That is, recognizing and conforming to group norms is adaptive for the self (e.g., leading to social acceptance and opportunities for collaborative learning) and the group (e.g., leading to more efficient group functioning and strength against competing groups; Tomasello, 2016). Thus, an individual who can quickly learn, adopt, and enforce their group's norms is a model group member as they increase their own (and their group's) functioning, resources, and survival.

If normative reasoning is indeed evolutionarily rooted and therefore a central and fundamental aspect of human cognition (Tomasello, 2016), children's descriptive-to-prescriptive tendency, even under novel and minimal conditions, should emerge cross-culturally. To date, however, this is unclear, as Roberts et al. (2017a, 2017b) conducted their research with U.S. samples. Indeed, the overreliance on samples recruited from WEIRD cultures (Western, Educated, Industrialized, Rich, and Democratic) is a problem for the general field of psychology (Henrich, Heine, & Norenzayan, 2010), the subfield of developmental psychology (Legare & Harris, 2016), and the specific research on children's normative reasoning, which includes mostly data derived from Germany, the U.S., or Sweden (for examples, see Cooley & Killen, 2015; Göckeritz et al., 2014; Josephs et al., 2016; Kenward, 2012; Roberts et al., 2017a, 2017b; Samland et al., 2016; Schmidt et al., 2016; but see research on "over-imitation," reviewed below, for a notable exception). Because of this, our theoretical models are founded upon the assumption that WEIRD cultures provide a standard from which to understand normative reasoning.

In an effort to test the extent to which children's descriptive-to-prescriptive tendency, with regard to novel groups engaged in innocuous behaviors, replicates (and varies) cross-culturally, and to move toward more inclusive, diverse, and generalizable data, we examined whether children recruited in mainland China, like children recruited in the U.S., interpreted descriptive group regularities as having normative force. Our focus on Chinese participants was motivated by previous research demonstrating that Chinese children, compared to U.S. children, engage in more group-based reasoning (as reviewed below), suggesting that Chinese children may also more often use descriptive group regularities to make prescriptive judgments. Thus, comparing children's descriptive-to-prescriptive tendency across two distinct cultures (i.e., U.S. and China) permitted a test of the replicability of previously detected effects with U.S. children (Roberts et al., 2017a, 2017b), while also testing for cross-cultural variation.

As discussed above, because normative reasoning may have an evolutionary basis (Tomasello, 2016), and because children quickly and intuitively enforce normative conformity (Rakoczy & Schmidt, 2013), it is possible that Chinese children, like U.S. children, would show a descriptive-to-prescriptive tendency. Further evidence for this possibility is cross-cultural similarity in children's tendency to "over-imitate" (i.e., faithfully reproduce modeled actions), suggesting that children's adherence to norms is a universal component of human cognition (see Nielsen & Tomaselli, 2010). Specifically, children across several diverse contexts (e.g., Aboriginal and Western Australia, Botswana, England, South Africa, Sweden) interpret the actions of adults as normative, imitating them precisely, even if they understand that the actions are inconsequential for achieving a goal (e.g., tapping a stick on the outside of a box before using it to dislodge a bolt; Horner & Whiten, 2005; Kenward, 2012; Nielsen & Mushin, 2014). Because children across diverse cultural contexts interpret neutral behaviors as normative in their

imitations, they may also interpret neutral descriptive group regularities as normative. Critically, though, the extent to which children engage in normative reasoning, including the extent to which they engage in imitative behavior, varies cross-culturally. For example, children in Vanuatu, a Pacific island nation, compared to children in the U.S., show higher rates of imitation (Clegg & Legare, 2016, see also Berl & Hewlett, 2015). Clegg and Legare (2016) argue that this cross-cultural variation could stem from non-Western cultures placing greater value on social conformity (see also Little, Carver, & Legare, 2016; Strachan, Samuel, & Takaro, 2007).

Indeed, in their seminal review, Markus and Kitayama (1991) discussed how people in Eastern cultures, such as China, tend to hold an interdependent view of the self, leading them to value social responsibility, interconnectedness, group solidarity, and fitting in. In contrast, people in Western cultures, such as the U.S., tend to hold an independent view of the self, which entails valuing self-expression, autonomy, uniqueness, and standing out (see also Henrich et al., 2010). There is of course variation in the degree to which individuals within cultures hold certain ideologies (e.g., not every person in the U.S. values independence over interdependence), though at the aggregate level, these cultural differences are well documented (see Oyserman, Coon, & Kemmelmeier, 2002), and they have important implications for behavior and social cognition. For example, Wang, Leichtman, and Davies (2000) documented that U.S. and Chinese families provided their children with different frameworks from which to create and understand narratives: storytelling within U.S. families typically involved reference to individual feelings, thoughts, and opinions, whereas storytelling within Chinese families more often involved reference to social relationships, rules, and responsibilities. Similarly, Miller, Wiley, Fung, and Liang (1997) found that storytelling within U.S. families involved reference to past positive behavior, which parents used for entertainment and affirmation purposes, whereas storytelling

within Taiwanese families involved reference to past transgressions, which parents used for moral socialization purposes. Cultural differences such as these have important implications for children's social cognition (Wang, 2016). For example, Li and Wang (2004) found that U.S. children believed that learning improved the *individual* (e.g., self-growth), whereas Chinese children believed that learning improved the larger *collective* (e.g., moral development). Relatedly, Wang (2004) asked children to recollect an early memory and to describe themselves and found that U.S. children tended to reference personal emotions and opinions, whereas Chinese children tended to reference social interactions and group activities (for comparable findings from adult samples, see Wang, 2001). In other work, Fu, Xu, Cameron, Heyman, and Lee (2007) found that Chinese children (compared to Canadian children) were more likely to disadvantage an individual to help a group than to disadvantage a group to help an individual (Fu et al., 2007), suggesting that they valued group needs over individual needs. Thus, because culture contributes to differences in children's behavior and social cognition, it may also contribute to differences in children's descriptive-to-prescriptive tendency.

Given the cross-cultural differences reviewed above, we were interested in the extent to which children's descriptive tendency replicated and varied cross-culturally. We examined whether Chinese children (ages 4 to 13), like same-aged U.S. children, used descriptive group regularities to make prescriptive judgments. As in Roberts et al. (2017a, 2017b), we introduced children to novel groups: Hibbles and Glerks, who were characterized in terms of morally neutral behaviors (e.g., the kind of music they listened to), and then to a series of conforming and non-conforming individuals (e.g., a Hibble who listened to music more typical of Glerks). We then measured children's *disapproval* of the behavior, *negativity* toward disapproved behaviors, and open-ended *explanations* regarding how they justified their evaluations. Including open-ended

responses was important, as they can be a more culturally sensitive method for detecting variation in children's concepts; they permit participants to share their own perspectives, use their own vocabularies, and generate their own responses (Hart & Edelman, 1992). We focused on 4- to 13-year-olds because this permitted a direct comparison to the age range in Roberts et al. (2017a), and because important developmental changes involving social concepts occur throughout these years (Bigler & Liben, 2006; Quintana, 1998; Rhodes & Gelman, 2009; Roberts & Gelman, 2015).

I had several hypotheses. First, because a preference for normative conformity may be evolutionarily rooted (Tomasello, 2016), and because early on in development children often take a strong normative stance (Rakoczy & Schmidt, 2013), we hypothesized that Chinese children, like U.S. children in previous research, would use descriptive group regularities (what *is*) to generate prescriptive judgments (what *should be*). That is, we predicted they would disapprove of non-conformity, evaluate it negatively, and use norm-based reasoning when explaining their evaluations (H1). Second, because Chinese children, like U.S. children, become increasingly likely with age to consider individual traits and attributes in their social reasoning (Wang, 2004), we hypothesized that with age, Chinese children would be more approving of non-conforming individuals (H2). Notably, however, an alternative hypothesis was that by virtue of having greater experience with socialization and enculturation of the concept of interconnectedness, with age, Chinese children would be more *disapproving* of non-conforming individuals. Third, because, Chinese culture, on average, is more likely than U.S. culture to value interdependence, we hypothesized that Chinese participants would be more likely than U.S. participants to show a descriptive-to-prescriptive tendency. Alternatively, because of the recent Westernization of Chinese culture (e.g., economic development, urbanization, increases in

formal education; see Pingali, 2006; Wu, 2009; Zeng & Greenfield, 2015), there may be no differences between Chinese and U.S. participants. To test for cross-cultural differences, we compared the data from the present research, which were collected in China, to the data collected by Roberts et al. (2017a), which were collected in the Midwestern U.S.

Method

Participants

In the People's Republic of China, four age groups of participants were included ($N = 172$): thirty-six 4- to 6-year-olds (39% female, M age = 4 years and 8 months, range = 4 years and 1 month to 6 years and 10 months), thirty-seven 7- to 9-year-olds (51% female, M age = 8 years and 5 months, range = 7 years and 2 months to 9 years and 11 months), forty-eight 10- to 13-year-olds (56% female, M age = 11 years, range = 10 years to 13 years and 9 months), and fifty-one adults (49% female, M age = 24, range = 18-40). All participants were recruited and tested by the second author. Children of the 4-6 age group were recruited and tested in a kindergarten, and children of other age groups were recruited and tested in an elementary school, both of which were located in a southwestern county in the Hubei Province of Central China, which has a population of approximately 500,000 residents; the Hubei Province is demographically 96% ethnic Han Chinese. Non-parent adults were recruited and tested at local cafés in a metropolis near the city as well as through word-of-mouth. In school, all children spoke Standard Mandarin Chinese, but at home, most of them spoke a local dialect of Mandarin Chinese.

Materials and procedure

The method was adapted from previous research (Roberts et al., 2017a) and presented via Qualtrics. The materials used in Roberts et al. (2017a) were translated from English into Chinese

by the second author, and then back-translated by one bilingual researcher to verify accuracy. First, participants were introduced to cartoon drawings of novel groups, “希博” (i.e., Hibbles) and “格里克” (i.e., 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.” Group membership was portrayed by clothing pattern (green stripes, orange triangles) and category labels (Hibbles, Glerks), and each group consisted of three individuals located on one side of the screen (left or right). Following the introduction of the novel groups, participants were given eight test trials presented in random order. Half of the trials depicted a non-conforming individual and the other half depicted a conforming individual. Across the eight trials, there were four behavioral domains: Music, Food, Games, and Language, with each behavior corresponding to the group in color (e.g., orange clothing pattern corresponded with an orange musical note, orange berries, an orange boomerang-shaped toy, and orange words in a foreign script within a speech bubble). On each trial, participants were given regularities for both groups and then shown a conforming or non-conforming individual. For example, “Hibbles listen to this kind of music (*pointing*) and Glerks listen to this kind of music (*pointing*). Look (*revealing the target*), this Hibble is listening to this [conforming or non-conforming] kind of music (*pointing*).” Participants were first asked to provide an *evaluation* (e.g., “Is it okay or not okay for this Hibble to listen to this kind of music?”). Participants who evaluated a behavior as “not okay” were then asked to provide a *negativity rating*; they 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*). Lastly, participants were asked to provide an *explanation* of their evaluation (e.g., why is it okay/not okay for this Hibble to listen to this kind of music?”). For an example trial in Chinese and in the English translation, please see Appendix B.

Measures and coding

First, we calculated the frequency with which disapproval (“not okay” evaluations) was given for both the conformity and non-conformity trials (potential range for each was 0 to 4). We focused on the frequency of “not okay” evaluations as the dependent variable, which reflected disapproval, because disapproval was our primary interest and because “okay” and “not okay” evaluations were precise inverses. Second, we calculated the average negativity ratings of participants who evaluated a behavior as “not okay” and were presented with the three-point negativity scale (potential range was 1 to 3). Third, participants' explanations were recorded verbatim in Chinese, translated into English by the second author, and coded by two research assistants who were blind to the hypotheses of the study (Cohen's kappa = .81). Disagreements were resolved by discussion.

Is in the previous studies, responses were coded into five types: (a) norm-based, (b) group-based, (c) individual-based, (d) similarity-based, and (e) other. See Table 1 for a description of the coding scheme. Codes were not mutually exclusive, thereby allowing participants to appeal to multiple explanations within a single response. We then examined within each response type (e.g., approved non-conformity) how often participants provided each particular explanation type (relative to the number of trials a particular response type was given). For example, a child who approved of non-conformity on 2 trials and gave an individual-based explanation on 1 of those trials was coded as having given individual-based explanations for approved non-conformity 50% of the time.

Parent survey. For exploratory purposes, parents and adult participants were also given the option of completing a survey (94% response rate) adapted from Feldman (2003) that assessed their views on conformity ($\alpha = .54$), respect for common norms ($\alpha = .51$), and

authoritarianism ($\alpha = .35$). Because these measure yielded low reliabilities and were not significantly related with participants' responses, they were not included in the primary analyses. However, we did test whether Chinese parents differed from a sample of U.S. parents in their value for conformity, respect for norms, and authoritarianism, and found that the Chinese sample tended to value group norms significantly more highly, consistent with previous research (Markus & Kitayama, 1991; see Table 11 for these data and analyses).

Results

The data were analyzed as they were Chapter 1, Study 1.

Disapproval toward conformity and non-conformity

There were significant main effects of age group, $F(3, 168) = 39.63, p < .001, \eta_p^2 = .41$, and behavior, $F(1, 168) = 346.13, p < .001, \eta_p^2 = .67$, and a significant interaction of age group and behavior, $F(3, 168) = 39.20, p < .001, \eta_p^2 = .41$. Consistent with our first hypothesis, all age groups were significantly more disapproving of non-conformity than of conformity ($ps \leq .009$), and consistent with our second hypothesis, disapproval toward non-conformity decreased at each age group ($ps \leq .032$; though the difference between 4- to 6-year-olds and 7- to 9-year-olds was non-significant, $p = .059$). Please see Tables 11 and 12 for all data and statistics, and Table 13 for the non-parametric data.

Negativity toward non-conformity

Focusing only on participants who disapproved of non-conformity and were subsequently asked how bad the behavior was (thirty-five 4- to 6-year-olds, thirty 7- to 9-year-olds, twenty-nine 10- to 13-year-olds, and seventeen adults), we found a main effect of age group, $F(3, 107) = 10.38, p < .001, \eta_p^2 = .23$. Both 4- to 6-year-olds ($M = 2.19, SE = .11$) and 7- to 9-year-olds ($M = 2.01, SE = .12$) were significantly more negative than 10- to 13-year-olds ($M = 1.41, SE = .12$)

and adults ($M = 1.49$, $SE = .16$) ($ps \leq .05$). The two youngest age groups did not differ significantly, nor did the two oldest age groups ($ps = 1.00$).

Explanations

See Table 14 for the explanation data.

Explanations about disapproved non-conformity

Focusing on participants who disapproved of non-conformity ($n = 111$), there were significant main effects of explanation type, $F(2.60, 278.30) = 27.52$, $p < .001$, $\eta_p^2 = .21$, and age group, $F(3, 107) = 8.32$, $p < .001$, $\eta_p^2 = .19$, and a significant interaction of explanation and age group, $F(7.79, 8.89) = 3.56$, $p = .001$, $\eta_p^2 = .09$. Four- to -six-year-olds gave mostly similarity-based and norm-based explanations, 7- to 9-year-olds and 10- to 13-year-olds gave mostly similarity-based and group-based explanations, and adults gave mostly group-based explanations. Together, these data indicate that when disapproving of non-conformity, younger children appealed to normative explanations (consistent with H1), whereas older children focused more on group-based explanations.

Explanations about approved non-conformity

Few 4- to 6-year-olds ($n = 2$) and 7- to 9-year-olds ($n = 10$) approved of non-conformity, so these age groups were excluded from this analysis. Thus, focusing on 10- to 13-year-olds ($n = 38$) and adults ($n = 46$) who approved of non-conformity, there were significant main effects of explanation type, $F(2.42, 198.08) = 42.80$, $p < .001$, $\eta_p^2 = .34$, and age group, $F(1, 82) = 6.02$, $p = .016$, $\eta_p^2 = .07$. Overall, both age groups gave mostly individual-based explanations when approving of non-conformity, with 10- to 13-year-olds overall giving more of each explanation type than adults (with the exception of similarity-based explanations).

Explanations about approved conformity

Focusing on participants who approved of conformity ($n = 171$), there was a main effect of explanation type, $F(2.52, 420.32) = 51.38, p < .001, \eta_p^2 = .24$, and age group, $F(3, 167) = 10.21, p < .001, \eta_p^2 = .16$, and a significant interaction of explanation type and age group, $F(7.55, 420.32) = 9.25, p < .001, \eta_p^2 = .14$. Children gave mostly group-based explanations, whereas adults gave mostly group-based, individual-based, and similarity-based explanations.

Secondary data analysis (China vs. U.S.)

We next compared the data derived from the present research with those derived from Roberts et al. (2017a, Study 1). The data analysis plan paralleled that of the main study, with the addition of “country” as a between-subjects variable. We report only the effects involving country, as cross-cultural differences were our a priori interest.

Disapproval toward conformity and non-conformity

We found main effects of behavior and age group, a significant interaction of behavior and country, $F(1, 272) = 6.33, p = .012, \eta_p^2 = .02$, and a significant interaction of age group, behavior, and country, $F(3, 272) = 6.53, p < .001, \eta_p^2 = .07$. Consistent with our third hypothesis (H3), Chinese 4- to 6-year-olds rated non-conformity more disapprovingly than same-aged U.S. children ($p = .042$), and they rated conformity less disapprovingly than same-aged U.S. children ($p < .001$). Rates of disapproval did not differ between countries at any other age group for ratings of non-conformity or conformity ($ps > .072$). The data are depicted graphically in Figure 10.

Negativity toward non-conformity

We next focused only on children ($n = 156$) who disapproved of non-conformity and were subsequently asked how bad the behavior was (only 4 U.S. adults disapproved of non-

conformity, so we excluded adults from this analysis), and did not find any significant effects involving country ($ps > .10$). Interestingly, this suggests that although Chinese children were more disapproving than U.S. children, those who were disapproving were not more negative.

Explanations about disapproved non-conformity

Focusing on children who disapproved of non-conformity ($n = 156$), we found a significant effect of country, $F(2, 150) = 30.54, p < .001, \eta = .07$, and a significant interaction between explanation type and country, $F(2.66, 399.65) = 5.33, p = .002, \eta = .03$. Consistent with our third prediction, overall, Chinese children provided more norm-based and similarity-based explanations than U.S. children ($ps < .02$).

Explanations about approved non-conformity. Because younger Chinese children rarely approved of non-conformity, we focused only on 10- to 13-year-olds and adults ($n = 127$), though this analysis yielded no significant effects on the basis of country ($ps > .072$).

Explanations about approved conformity. Focusing on children who approved of conformity ($n = 275$), we found a significant effect of country, $F(1, 267) = 11.85, p = .001, \eta < .04$, a significant interaction of age group and country, $F(1, 267) = 6.77, p < .001, \eta = .07$, a significant interaction of explanation type and country, $F(2.75, 733.01) = 11.20, p < .001, \eta = .04$, and a significant interaction of explanation type, age group, and country, $F(8.23, 733.01) = 3.35, p = .001, \eta = .04$. All Chinese participant groups gave more similarity-based explanations than U.S. participant groups ($ps \leq .042$). Chinese 4- to 6-year-olds gave more group-based explanations, but fewer norm-based explanations, compared to same-aged U.S. children ($ps \leq .004$).

Discussion

The present results indicate that children's descriptive-to-prescriptive tendency replicates cross-culturally. More specifically, consistent with research with same-aged U.S. children (Roberts et al., 2017a), and consistent with our first hypothesis, Chinese children disapproved of non-conformity, rated it negatively, and provided norm-based explanations when justifying their evaluations (e.g., if Hibbles were characterized as listening to a specific kind of music, individual Hibbles *should* listen to that kind of music, and it is bad if they did not). These data are consistent with the proposal that a preference for normative conformity is deeply rooted in our evolutionary history (Chudek & Henrich, 2011; Tomasello, 2016), likely as a result of the central role that group-based living played throughout human history (e.g., conforming to group norms enhanced the functionality, resources, and survival of the self and the group). Also, again replicating previous research, Chinese children's descriptive-to-prescriptive tendency declined at each age group (i.e., regarding non-conformity, the youngest age groups were most likely to disapprove, show negativity, and provide norm-based explanations), which aligns with our second hypothesis and previous research suggesting that Chinese children (like U.S. children) become more sensitive to individual differences with age (Wang, 2004). Of course, additional work is needed to examine these questions with a broader range of cultural contexts, but for now, they provide empirical evidence that children from two different cultural contexts interpret descriptive group regularities as having prescriptive force.

Critically, though, the strength of this tendency varied cross-culturally (consistent with our third hypothesis). Compared to same-aged U.S. children, Chinese 4- to 6-year-olds evaluated non-conformity more disapprovingly (and conformity more approvingly), and across all child groups, Chinese children provided more norm-based explanations when justifying their

disapproval. Additionally, unlike U.S. adults in previous research, Chinese adults in the present research were more disapproving of non-conformity than of conformity. This finding was especially interesting given that the task was explicit and involved novel groups engaged in morally neutral behaviors, suggesting that even under such circumstances, Chinese adults show a descriptive-to prescriptive tendency. Taken together, these data are consistent with the reasoning that individuals from cultures that more highly value interdependence may be more likely to use group regularities to disapprove of non-conformity (because doing so privileges group functioning over individual functioning). Thus, the present data provide evidence that although the descriptive-to-prescriptive tendency replicates across cultures, the degree of it may not.

Of course, additional research is needed to understand precisely which cultural factors contributed to the cultural variation detected here. As mentioned previously, we reasoned that these factors could have included larger historical and societal values (Markus & Kitayama, 1991; Oyserman et al., 2012) or parental socialization (Miller et al., 1997; Wang, 2016). Indeed, our parental survey, although it was itself unreliable, demonstrated that Chinese parents placed greater value on conformity, respect for norms, and authoritarianism than U.S. parents. Thus, Chinese children may indeed be taught, explicitly, to place greater value on group norms, which may directly influence their descriptive-to-prescriptive tendency. I look forward to future research designed to test parent-based pedagogy more systematically.

One alternative explanation of our findings involves a pedagogical interpretation of the task: a knowledgeable experimenter conveyed group regularities and thus children may have believed that the experimenter conveyed information for the child's pedagogical benefit (see Butler & Markman, 2012). If so, it would be reasonable for children to disapprove of a non-conforming individual. Although possible, we argue that pedagogy does not completely explain

children's descriptive-to-prescriptive tendency, as measured here. First, past work with the same methodology shows that when the experimenter describes a single individual engaging in a behavior, rather than groups, children are more accepting of non-conformity (see Roberts et al., 2017a, Study 3; Roberts et al., 2017b, control condition). If children were simply interpreting a given behavior as appropriate because they infer that the experimenter is describing it to them for pedagogical purposes, then it should not matter whether an individual or a group performed the behavior. Second, Chinese children were more prescriptive in their responses than were U.S. children, suggesting that larger cultural values, rather than pedagogy per se, influenced children's descriptive-to-prescriptive tendency.

An important question for future research is to explore precisely why prescriptiveness declined with age, which was especially notable given the literature suggesting that there is greater value placed on interdependence in China than in the U.S. (e.g., Markus & Kitayama, 1991), and given the alternative prediction that children in more interdependent cultures would be increasingly prescriptive with age as a function of their socialization experiences. One possibility is that older Chinese children were better than their younger peers at suppressing and overriding their responses on this explicit task (see also the discussion in Chapter 5). Indeed, U.S. data on racial attitudes suggests that whereas explicit biases decline with age, implicit biases often persist (e.g., Apfelbaum, Pauker, Ambady, Sommers, & Norton, 2008; Baron & Banaji, 2006). Thus, future research should test whether older children and adults make prescriptive judgments at an implicit level. A second possibility is that because the present task provided a particularly strong test of attention to conformity (given the novel categories and innocuous behaviors), older children and adults may have required additional information or more consequential behaviors (e.g., where failure to conform has more overt consequences for

the group). Under conditions with more severe acts of non-conformity, older children and adults may indeed show a descriptive-to-prescriptive tendency, even on an explicit task such as this. A third possibility is that older children were more aware of individual differences, thereby making them more likely to reason that different people simply like different things (see Quintana, 1998). The explanation data align with this last interpretation: When 10- to 13-year-olds approved of non-conformity, they most often appealed to individual-based explanations (e.g., “Everyone has their own thoughts [and] playing with a toy is part of one’s own thoughts. There should be no limit on that”). If this was indeed the case, future research should test whether encouraging children, especially younger children, to think about individual differences reduces their descriptive-to-prescriptive tendency.

Importantly, I acknowledge that there exists much variation within cultures. Certainly, a variety of factors are likely to influence a child's attention to group norms, whether in the U.S. or China (e.g., rural vs. urban, liberal vs. conservative, part of a majority or minority) and thus their acceptance of non-conformity (for data on variation in social group concepts, see Rhodes & Gelman, 2009 for context-based variation, and Roberts & Gelman, 2016 for group membership-based variation). We look forward to future research that explores within-culture variation more systematically, as doing so would provide a richer and more nuanced understanding of the effects detected here. (See also Talhelm et al., 2014 for a description of differences in interdependence within China as a function of rice vs. wheat farming. Notably, the Hubei Province studied here is one of the provinces that Talhelm et al. (2014) identified as relatively interdependent.) For now, however, the present research documents that children’s descriptive-to-prescriptive tendency both replicates and varies cross-culturally.

Lastly, an important issue is the extent to which the present findings with novel groups to which children did not belong would extend to real-life settings. Certainly, children have learned to associate stereotypical features with a variety of social groups. For example, English children report that English soccer fans are friendly (Abrams et al., 2009), U.S. children report that boys play with toy trucks (Taylor, Rhodes, & Gelman, 2009), Israeli children report that Arabic people wear scarves (Barrett, 1996), and South African children report that white people are wealthy (Olson, Shutts, Kinzler, & Weisman, 2012). Children's descriptive-to-prescriptive tendency, augmented by years of real-world experiences (e.g., socialization, group membership, moral concerns), may lead to the inference that English soccer fans *should be* friendly, boys *should* play with toy trucks, Arabic people *should* wear scarves, and white people *should be* wealthy. Thus, the early emergence and pervasiveness of stereotype development may be firmly rooted in children's reasoning about social group norms, and data presented in Chapter four suggest that this reasoning replicates (and varies) cross-culturally.

Table 11

Parent Survey (Chapter IV). Mean frequencies and standard errors from the parent survey (adult self-report data). U.S. data come from Roberts et al. (2017b; Chapter III).

			M(SE)		
Country	Age	N	Conformity	Respect for Norms	Authoritarian
China	4-6	39	4.05(.10)	3.56(.11)	2.18(.18)
	7-9	34	4.07(.10)	3.24(.12)	2.24(.19)
	10-13	42	3.93(.07)	3.08(.08)	2.12(.14)
	Adult	50	3.52(.10)	2.74(.11)	1.30(.19)
U.S.	4-6	107	3.04(.06)	2.66(.07)	1.08(.11)
	7-9	78	2.94(.07)	2.76(.08)	1.33(.13)
	10-13	62	3.58(.08)	3.08(.10)	1.57(.14)
	Adult	n/a	n/a	n/a	n/a

Note. Scores could range from 0 to 6. Although the parent survey was unreliable and uncorrelated with children's responses, we tested whether Chinese parents differed from U.S. parents in their value for conformity, respect for norms, and authoritarianism. Data from the U.S. participants were from Roberts et al. (2017b), which included participants from the same source as those recruited in Roberts et al. (2017a). We did not include data from the adult participants because of insufficient data from U.S. adults. A 2 (country: China, U.S.) by 2 (age group: 4-6, 7-9) multivariate ANOVA with country and age group as between-subjects variables, and value for conformity, respect for common norms, and authoritarianism as the dependent variables, yielded main effects of country on all three outcomes: Conformity, $F(1, 356) = 135.82, p < .001, \eta = .28$, respect for norms, $F(1, 356) = 31.00, p < .001, \eta = .08$, and authoritarianism, $F(1, 356) = 47.47, p < .001, \eta = .12$. On each, Chinese parents gave higher ratings than U.S. parents. There was also an interaction between country and age group on conformity, $F(2, 362) = 11.69, p < .001, \eta = .06$, and on respect for norms, $F(2, 362) = 6.20, p = .002, \eta = .03$. Pairwise comparisons showed that on each outcome, Chinese parents gave higher scores than did U.S. parents (all $ps < .02$; though there were no country-based differences with regard to respect for norms for 10- to 13-year-olds, $p = .59$).

Table 12.

Disapproval data (Chapter IV). Means, standard errors, and one-sample *t*-test statistics comparing average disapproval frequencies against chance (i.e., 2) across age groups and behavior type.

Age	Behavior	M(SE)	<i>t</i>	<i>p</i>	<i>d</i>
4-6	Non-conformity	3.86(.11)	16.36	< .001	2.72
	Conformity	.17(.09)	-19.62	< .001	3.27
7-9	Non-conformity	3.03(.27)	3.85	< .001	.63
	Conformity	.14(.09)	21.18	< .001	3.45
10-13	Non-conformity	1.50(.23)	-2.17	.035	.33
	Conformity	.13(.05)	-38.87	< .001	5.61
Adults	Non-conformity	.73(.18)	-7.10	< .001	.99
	Conformity	.18(.10)	-18.27	< .001	2.56

Table 13.

Non-parametric data (Chapter IV). Non-parametric Wilcoxon signed-rank tests across age groups and behaviors, indicating how many Chinese children most often approved (okay), most often disapproved (not okay), or approved and disapproved equally (Tie).

Age	Behavior	OK	Not OK	Tie	Z	<i>p</i>	<i>r</i>
4-6	Non-conformity	1	35	0	5.84	<.001	.97
	Conformity	32	4	0	5.01	<.001	.84
7-9	Non-conformity	7	30	0	3.31	< .001	.55
	Conformity	34	3	0	5.21	< .001	.86
10-13	Non-conformity	29	19	0	1.96	.005	.28
	Conformity	42	6	0	6.36	<.001	.92
Adults	Non-conformity	34	17	0	1.99	.047	.29
	Conformity	47	4	0	5.82	< .001	.82

Table 14.

Explanation data (Chapter IV). Percentage of explanation types for each behavior, across evaluation types and age groups.

Percentage of Explanation Types							
M(SE)							
Behavior	Evaluation	Age	N	Norm	Group	Individual	Similarity
Non-Conformity	Not Okay	4-6	35	57.1(5.4)	45(7.4)	0	64.3(7.2)
		7-9	30	23.6(5.9)	52.5(8.0)	5.8(3.9)	44.2(7.8)
		10-13	29	20.4(6.0)	59.2(8.2)	14.1(4.0)	44.3(7.9)
		Adult	17	4.4(7.8)	42.6(10.7)	5.9(5.2)	23.5(10.4)
Non-conformity	Okay	4-6	2	12.5(12.2)	25.0(21.2)	12.5(24.4)	12.5(10.5)
		7-9	10	2.5(5.5)	28.3(9.5)	70.8(10.9)	13.3(4.7)
		10-13	38	11.6(2.8)	23.7(4.9)	54.4(5.6)	3.3(2.4)
		Adult	46	4.3(2.6)	15.8(4.4)	43.1(5.1)	8.3(2.2)
Conformity	Okay	4-6	36	3.0(2.2)	51.6(6.3)	4.2(4.7)	71.5(6.3)
		7-9	37	6.1(2.2)	57.2(6.2)	12.2(4.6)	40.8(6.2)
		10-13	48	8.2(1.9)	42.9(5.5)	28.3(4.0)	27.8(5.4)
		Adult	50	5.5(1.9)	23.0(5.4)	26.0(4.0)	20.5(5.3)

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 codes are not reported (thus explaining why the percentages add to more than or less than 100). Because participants rarely disapproved of conformity, data for this response type are not presented.

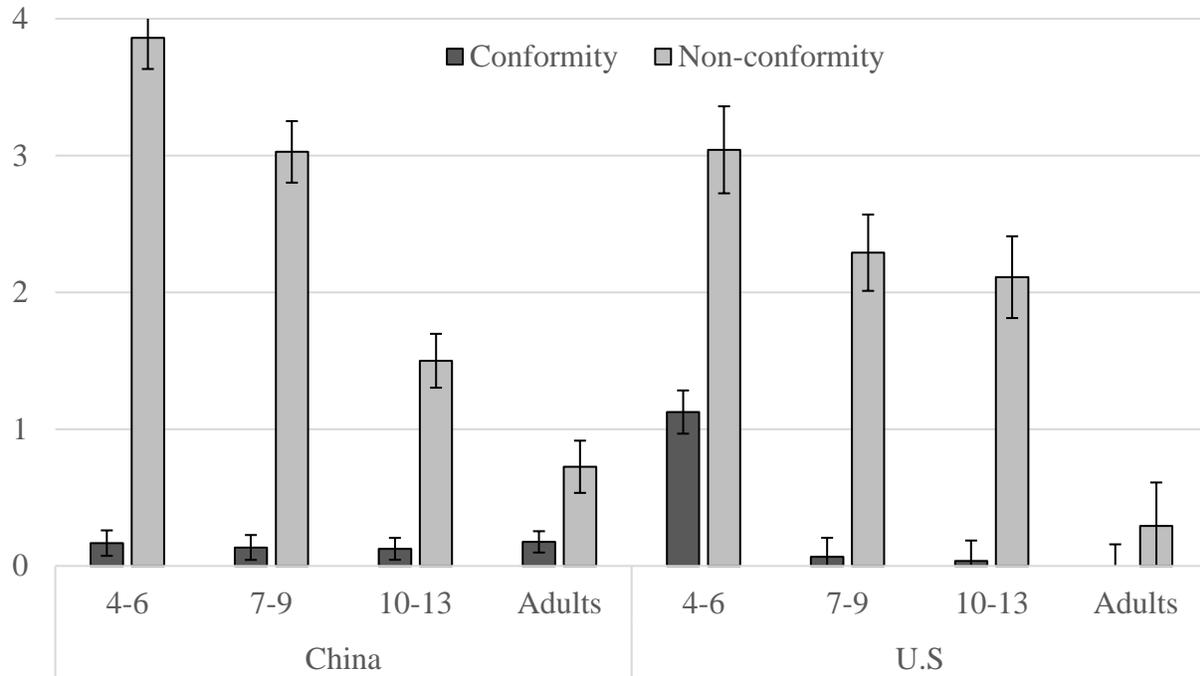


Figure 10. Disapproval data (Chapter IV). Mean frequency of disapproval of conformity and non-conformity across age groups and country. Scores could range from 0 to 4. Bars depict standard errors.

CHAPTER V

General Discussion

In the studies presented in this dissertation, I examined children's and adults' use of descriptive group regularities to make prescriptive judgments. In Chapter 2, I found that children interpreted the regularities as prescriptive, even when the groups were novel and third-party, and the behaviors were innocuous. I also found that this descriptive-to-prescriptive tendency was robust across intergroup contexts (i.e., regardless of whether a group is in competition or cooperation with another, individual group members should conform to group norms), and that an emphasis on individuality reduced children's prescriptiveness. In Chapter 3, I tested how visual and linguistic cues contributed to this tendency, and found that prescriptive judgments were elicited when the group regularities were conveyed through group presence, category labels, or generic statements (and not in the absence of these factors). In Chapter 4, I examined whether children's descriptive-to-prescriptive tendency replicated internationally, and found that the tendency to use group regularities to make prescriptive judgments emerged in early childhood in two distinct cultures, though cultural values nonetheless influenced the degree to which it predominates across these two populations. Although these findings do not generalize beyond these two populations, they provide preliminary evidence for a cross-cultural descriptive-to-prescriptive tendency.

Overall, these data have important implications for furthering the understanding of children's social reasoning, and they provide empirical support for an early emerging descriptive-to-prescriptive tendency that persists even in the absence of additional factors

previously shown to foster prescriptive judgments. That is, unlike previous research in developmental, social, and cognitive psychology in which children's prescriptive judgments were found to be fostered by socialization and familiarity with the groups (Bigler & Liben, 2006), in-group membership (Rutland et al., 2015; Tajfel & Turner, 1979), beliefs about what is morally appropriate (Cooley & Kileen, 2015; Smetana et al., 2012), or beliefs about between-group cooperation and allegiances (McAuliffe & Dunham, 2016; Rhodes, 2013; Sherif & Sherif, 1953), children in the present research prescriptively evaluated non-conformity even with respect to harmless behaviors of novel, third-party groups, and irrespective of whether those groups were in competition or cooperation with one another. Again, children's descriptive-to-prescriptive tendency was elicited by minimal input (Chapter 3), and it replicated across international contexts (Chapter 4). Together, these findings provide strong support for the notion that group norms play a foundational role in social cognition (Tomasello, 2016).

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.

Limitations and future directions

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, 2016). 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 eliciting such reasoning may change with age. Certainly, sociopolitical contexts and beliefs shape social perception (e.g., Dunham, 2011; Ho et al., 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 et al., 2008; Baron & Banaji, 2006; Dunham et al., 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. For example, older children and adults placed under time demands may show more intuitive, reflexive, and rapid judgments, making prescriptive judgments at a rate comparable to younger children.

Future research should include individual difference measures (e.g., psychological essentialism) to gain further insights, as children predisposed to attend more to group-based information may be more sensitive to group norms. 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 may have relied on notions of group loyalty, whereas those who conceptualized the groups as task focused may have relied on notions of interdependence (see Lickel et al., 2000; Plötner, Over, Carpenter, & Tomasello, 2016). Another future direction will be to examine the extent to which children's descriptive-to-prescriptive tendency influences their real-world behavior. In the present studies, I assessed children's beliefs and evaluations of what is ideal and acceptable (i.e., prescriptive judgments; Bear & Knobe, 2016 and not their active efforts to maintain and perpetuate norms (i.e., norm enforcement; Rakoczy & Schmidt, 2013). Do children also disapprove of non-conformity, explicitly or implicitly, when they encounter it in the real-world? Previous research suggests that they would. As mentioned previously, when children observe a tool being used in a certain way, they protest against individuals who use the tool differently (e.g., "You *must* do this"; Kenward, 2012). Similarly, when children learn the rules of a game and observe someone violate those rules, they respond with explicit protest and critique (e.g., "You *can't* do that"; Rakoczy & Schmidt, 2013). Moreover, children spontaneously create their own norms, teach them to others, and negatively evaluate those who do not conform (Göckeritz et al., 2014). Thus, it is possible that when children learn a group norm, even one that

is novel, third-party, and harmless, they spontaneously critique those who fail to conform.

Observational studies in which children are shown group norms and subsequently encounter a non-conforming individual would do well to determine the extent to which this tendency extends beyond more tightly controlled lab settings, and would speak more directly to real-world norm enforcement.

These dissertation data are also limited in that they speak only to group norms that involve behaviors (i.e., language spoken, music listened to, games played, food eaten). An open question is whether children also adhere to group norms that involve beliefs. That is, if a group is characterized by a particular belief, do children infer that individual group members should also hold that belief? Work conducted by Heiphetz, Spelke, Harris, and Banaji (2013) provided some preliminary insights. More specifically, they found that children were most likely to conceptualize a belief as objective (i.e., that the belief was a “right” or “correct” proposition) when that belief was factual (e.g., the sky is blue), followed by when that belief was ideological (e.g., that there is one God), followed by preference-based beliefs (e.g., that roses are beautiful). When reasoning about factual beliefs, children may maintain that it is appropriate to conform to what is objectively correct, irrespective of what the group believes (e.g., if the group believes that a blue sky is red, individual group members should not conform). When reasoning about preference-based beliefs, children may maintain that it is appropriate to hold any belief, irrespective of what the group believes (e.g., if the group thinks roses are beautiful, it is up to individual group members to conform or not). When reasoning about ideological beliefs, which again, are conceptualized as entailing an element of fact and preference (Heiphetz et al., 2013), it is unclear whether children believe that individual group members should conform to the ideological beliefs of the group (e.g., if the group believes in a specific deity, should individual

group members conform?). Systematically testing these questions would further the understanding of the domains in which children interpret group norms as prescriptive.

Conclusion

In conclusion, 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, and many additional questions are yet to be explored. For example, are these effects specific to the human domain, or do they hold even when reasoning about non-human animals (for research on domain-based category judgments, see Keil, 1989; Rhodes & Gelman, 2009)? 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; Rhodes, 2012; Smetana et al., 2012)? These important questions await future research, and promise to further reveal children's social-cognitive capacities.

Appendix A – Item Effects

In Chapter 2, additional analyses were conducted to ensure that the results did not vary by the specific item (i.e., berry, language, games, music). These data are presented below.

Study 1

Disapproval toward conformity and non-conformity. We conducted a 4 (age group: 4-6, 7-9, 10-13, adult) x 2 (behavior type: conformity, non-conformity) x 4 (domain: food, games, language, music) repeated measured ANOVA with age group as a between-subjects variable, behavior and domain as within-subjects variables, and the frequency of not-okay evaluations as the dependent variable. There was no main effect of domain, $F(3, 306) = .89, p = .45, \eta_p^2 = .009$, two way interaction of domain and age group, $F(9, 306) = .95, p = .49, \eta_p^2 = .03$, two way interaction effect of domain and behavior, $F(3, 306) = .41, p = .75, \eta_p^2 = .004$, or three way interaction of domain, age group, and behavior, $F(3, 306) = .87, p = .55, \eta_p^2 = .03$. These analyses were supported by a series of one-sample *t*-tests comparing disapproval rates to chance (i.e., 5) for each item (i.e., food, games, language, music). Consistent with the general patterns, for conformity, all participant groups were below chance on every item: 4 to 6 ($ps < .05$). For non-conformity, 4- to 6-year-olds were above chance on all items ($ps < .04$), 7- to 9-year-olds and 10- to 13-year-olds were at chance on all items ($ps > .21$), and adults were below chance on all items ($ps < .001$).

Negativity toward non-conformity. Next, we focused only on children who disapproved of non-conformity and were asked how bad the non-conforming behavior was (1 = *kind of bad*, 2 = *pretty bad*, 3 = *very, very bad*; $n = 64$). Again, we focused on children because only 1 adult

disapproved of non-conformity. Because not all children disapproved of all items (e.g., a child could have disapproved of non-conformity for music but not for language), we could not make a direct comparison across items. Therefore, we conducted four univariate ANOVAs with age group as the independent variable and negativity toward a specific item (i.e., berries, games, language, music) as the dependent variable. Overall, each analysis showed that within each domain, negativity was higher in the youngest age group compared to the oldest age group. Regarding berries, there was a marginally significant effect of age group, $F(2, 49) = 2.71, p = .077, \eta_p^2 = .10$, showing that 4- to 6-year-olds ($M = 2.42, SE = .20$) were marginally more negative than 10- to 13-year-olds ($M = 1.75, SE = .21$) ($p = .07$). For games, there was again a significant effect of age group, $F(2, 46) = 4.16, p = .022, \eta_p^2 = .15$, showing that 4- to 6-year-olds ($M = 2.42, SE = .19$) were more negative than 10- to 13-year-olds ($M = 1.62, SE = .23$) ($p = .03$). For language, there was a significant effect of age group, $F(2, 48) = 3.67, p = .03, \eta_p^2 = .13$, showing that 4- to 6-year-olds ($M = 2.53, SE = .19$) were more negative than 10- to 13-year-olds ($M = 1.80, SE = .20$) ($p = .028$). For music, there was a significant effect of age group, $F(2, 46) = 4.84, p = .012, \eta_p^2 = .17$, showing that 4- to 6-year-olds ($M = 2.44, SE = .19$) were more negative than 10- to 13-year-olds ($M = 1.54, SE = .22$) ($p = .01$).

Study 2

Disapproval toward conformity and non-conformity. We conducted a 4 (age group: 4-6, 7-9, 10-13, adult) x 2 (behavior: conformity, non-conformity) x 4 (domain: food, games, language, music) x 2 (condition: competition, cooperation) repeated measured ANOVA, with age group and condition as between-subjects variables, behavior and domain as within-subjects variables, and the frequency of not-okay evaluations as the dependent variable. There was no main effect of domain, $F(3, 558) = .85, p = .47, \eta_p^2 = .005$, two way interaction of domain and

age group, $F(9, 558) = 1.52, p = .14, \eta_p^2 = .02$, two way interaction of domain and behavior, $F(3, 558) = 1.49, p = .22, \eta_p^2 = .008$, two way interaction of domain and condition, $F(3, 558) = .16, p = .92, \eta_p^2 = .001$, domain, age group, and condition, $F(9, 558) = 1.69, p = .09, \eta_p^2 = .03$, three way interaction of domain, age group, and behavior, $F(9, 558) = 2.71, p = .077, \eta_p^2 = .10$, three way interaction of domain, behavior, and condition, $F(3, 5587) = .03, p = .99, \eta_p^2 < .001$, or a four way interaction of domain, behavior, age group, and condition, $F(9, 558) = .82, p = .60, \eta_p^2 = .013$. Also, consistent with the pattern presented in the main text, for conformity, all age groups were below chance on all items ($ps < .001$). For non-conformity, across all items, 4- to 6-year-olds and 7- to 9-year-olds were above chance ($ps \leq .003$), 10- to 13-year-olds were at chance ($ps > .56$), and adults were below chance ($ps \leq .001$).

Negativity toward non-conformity. Next, we again focused only on children (not adults because of insufficient data) who disapproved of non-conformity and were asked how bad the non-conforming behavior was. We again conducted four univariate ANOVAs with age group as the independent variable and negativity toward a specific item (i.e., berries, games, language, music) as the dependent variable. There were no significant age group effects regarding berries, $F(2, 94) = .76, p = .47, \eta_p^2 = .016$, language, $F(2, 94) = .68, p = .51, \eta_p^2 = .014$, or music, $F(2, 91) = .32, p = .73, \eta_p^2 = .007$. Notably, there was a significant age group effect regarding games, $F(2, 90) = 3.57, p = .03, \eta_p^2 = .074$. Pairwise comparisons showed that 4- to 6-year-olds ($M = 2.26, SE = .14$) were marginally more negative than 10- to 13-year-olds ($M = 1.71, SE = .18$) ($p = .053$).

Study 3

Disapproval toward conformity and non-conformity. One-sample t -tests to compared disapproval rates against chance (i.e., .5) for each domain, showing that for conformity,

disapproval was below chance across all domains (though marginally for berries): Berries: $M = .32$, $SE = .10$, $t(25) = -1.89$, $p = .07$; Games: $M = .20$, $SE = .08$, $t(25) = -3.67$, $p = .001$; Language: $M = .16$, $SE = .08$, $t(25) = -4.54$, $p < .001$; Music: $M = .20$, $SE = .08$, $t(25) = -3.67$, $p = .001$. For non-conformity, disapproval was at chance across all domains: Berries: $M = .36$, $SE = .10$, $t(25) = -1.43$, $p = .17$; Games: $M = .36$, $SE = .10$, $t(25) = -1.43$, $p = .16$; Language: $M = .40$, $SE = .10$, $t(25) = -1.00$, $p = .33$; Music: $M = .36$, $SE = .10$, $t(25) = -1.43$, $p = .16$. A series of independent samples t -test showed that for each domain, disapproval rates between conformity and non-conformity did not differ significantly from one another (though there was a marginal difference for berries, $p = .06$).

Negativity toward non-conformity. For all items, children's negativity toward non-conformity were in the "a little bad" range: Berries ($M = 2.33$, $SE = .27$), Games ($M = 2.11$, $SE = .29$), Language ($M = 1.90$, $SE = .26$), Music ($M = 2.11$, $SE = .29$). A series of paired-sample t -tests showed that these values did not differ significantly from one another ($p > .50$).

Appendix B – Chinese and English Scripts

Chinese (the administered language)

“下面我要给你介绍两个不同的小组。一个小组(指向)叫做希博，一个小组(指向)叫做格里克。”

“希博们(指向)吃这几种浆果，而格里克们(指向)吃这几种浆果。看，这只希博(指向目标人物)正在吃这几种浆果。”

评估：“这只希博吃这几种浆果行还是不行呢？”

负面：“[如果参与者认为不可以]“那么他这样做是有一点不好，很不好，还是非常非常不好呢？”

解释：“为什么这只格里克吃这几种浆果[行或不行]呢？”

English translation

“I’m going to tell you about two kinds of groups. This group (pointing) is called Hibbles, and this group (pointing) is called Glerks.”

“Hibbles (pointing) eat these kind of berries and Glerks (pointing) eat these kinds of berries. Look, this Glerk (pointing to target) is eating these kind of berries.”

Evaluation: “Is it okay or not okay for this Glerk to eat these kind of berries?”

Negativity: [if participant evaluated this as not okay] “Is it a little bad, pretty bad, or very, very bad, for this Hibble to eat these kind of berries?”

Explanation: “Why is it [okay or not okay] for this Glerk to eat these kind of berries?”

REFERENCES

- Abrams, D. (2011). Wherein lies children's intergroup bias? Egocentrism, social understanding, and social projection. *Child Development, 82*, 1579–1593. doi:10.1111/j.1467-8624.2011.01617.x
- Abrams, D., Rutland, A., & Cameron, L. (2003). The development of subjective group dynamics: Children's judgments of normative and deviant in-group and out-group individuals. *Child Development, 74*, 1840–1856. doi.org/10.1046/j.1467-8624.2003.00641.x
- Abrams, D., Rutland, A., Pelletier, J., & Ferrell, J. M. (2009). Children's group nous: Understanding and applying peer exclusion within and between groups. *Child Development, 80*, 224–243. doi:10.1111/j.1467-8624.2008.01256.x
- Allport, G. (1954). *The nature of prejudice*. Reading, MA: Addison-Wesley.
- Apfelbaum, E. P., Pauker, K., Ambady, N., Sommers, S. R., & Norton, M. I. (2008). Learning (not) to talk about race: When older children underperform in social categorization. *Developmental Psychology, 44*, 1513–1518. doi:10.1037/a0012835
- Baron, A. S., & Banaji, M. R. (2006). The development of implicit attitudes: Evidence of race evaluations from ages 6 and 10 and adulthood. *Psychological Science, 17*, 53–58. doi:10.1037/t03782-000
- Barrett, M. (1996). Development of social categories and stereotypes in early childhood: the case of “the Arab” concept formation, stereotype, and attitudes by Jewish children in Israel.

International Journal of Intercultural Relations, 20, 341–370. doi: 10.1016/0147-1767(96)00023-5

Bear, A., & Knobe, J. (2016). Normality: Part descriptive, part prescriptive. *Cognition*.

Advanced online publication. 10.1016/j.cognition.2016.10.024

Berl, R. E. W., & Hewlett, B. S. (2015). Cultural variation in the use of overimitation by the Aka and Nganda of the Congo Basin. *PLOS ONE*, 10, 1-20. doi:

10.1371/journal.pone.0120180

Bigler, R. S., Jones, L. C., & Lobliner, D. B. (1997). Social categorization and the formation of intergroup attitudes in children. *Child Development*, 68, 530–543. doi:10.2307/1131676

Bigler, R. S., & Liben, L. S. (2006). A developmental intergroup theory of social stereotypes and prejudice. In R. V. Kail (Ed.), *Advances in child development and behavior* (Vol. 34, pp. 39-89). San Diego, CA: Elsevier.

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

doi:10.1023/A:1023574427720

Casler, K., Bickel, L., & Hackett, E. (2013). 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*, 29, 2156–2160. doi:10.1016/j.chb.2013.05.009

Chudek, M., & Henrich, J. (2011). Culture-gene coevolution, norm-psychology and the emergence of human prosociality. *Trends in Cognitive Sciences*, 15, 218–226. doi:

10.1016/j.tics.2011.03.003

Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology*, 55, 591–621. doi:10.1146/annurev.psych.55.090902.142015

- Cimpian, A., & Erickson, L. C. (2012). Remembering kinds: New evidence that categories are privileged in children's thinking. *Cognitive Psychology*, *64*, 161–185. doi: 10.1016/j.cogpsych.2011.11.002
- Claidière, N., & Whiten, A. (2012). Integrating the study of conformity and culture in humans and nonhuman animals. *Psychological Bulletin*, *138*, 126–145. doi:10.1037/a0025868
- Clegg, J. M., & Legare, C. H. (2016). A cross-cultural comparison of children's imitative flexibility. *Developmental Psychology*, *52*, 1435-1444. doi: 10.1037/dev0000131
- Cooley, S., & Killen, M. (2015). Children's evaluations of resource allocation in the context of group norms. *Developmental Psychology*, *51*, 554–563. doi: 1037/a0038796
- Corriveau, K. H., Fusaro, M., & Harris, P. L. (2009). Going with the flow: Preschoolers prefer nondissenters as informants. *Psychological Science*, *20*, 372–377. doi:10.1111/j.1467-9280.2009.02291.x
- Cosmides, L., Tooby, J., & Kurzban, R. (2003). Perceptions of race. *Trends in Cognitive Sciences*, *7*, 173–179. doi:10.1016/S1364-6613(03)00057-3
- Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*, 5–18. doi:10.1037/0022-3514.56.1.5
- Diesendruck, G., & Menahem, R. (2015). Essentialism promotes children's inter-ethnic bias. *Frontiers in Psychology*, *6*, 1–8. doi:10.3389/fpsyg.2015.01180
- Dunham, Y. (2011). An angry = outgroup effect. *Journal of Experimental Social Psychology*, *47*, 668–671. doi:10.1016/j.jesp.2011.01.003
- Dunham, Y., Baron, A. S., & Banaji, M. R. (2008). The development of implicit intergroup cognition. *Trends in Cognitive Sciences*, *12*, 248–253. doi:10.1016/j.tics.2008.04.006

- Durkee, M. I., & Williams, J. L. (2013). Accusations of acting white: Links to black students' racial identity and mental health. *Journal of Black Psychology, 41*, 26–48.
doi:10.1177/0095798413505323
- Eidsen, C. R., & Coley, J. D. (2014). Not so fast: Reassessing gender essentialism in young adults. *Journal of Cognition and Development, 15*, 382–392.
- Engelmann, J. M., Herrmann, E., Rapp, D. J., & Tomasello, M. (2016). Young children (sometimes) do the right thing even when their peers do not. *Cognitive Development, 39*, 86–92. doi: 10.1016/j.cogdev.2016.04.004
- Feldman, S. (2003). Social conformity: A theory of enforcing authoritarianism. *Political Psychology, 24*, 41-74. doi: 10.1111/0162-895X.00316
- Field, A. (2011). *Discovering statistics using SPSS* (3rd ed.). Thousand Oaks, CA: SAGE Publications.
- Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology, 82*, 878–902. doi: 10.1037//0022-3514.82.6.878
- Fiske, S. T., & Neuberg, S. L. (1990). A continuum model of impression formation: From category based to individuating processes as a function of information, motivation, and attention. *Advanced in Experimental Social Psychology, 23*, 1–74. doi:10.1016/S0065-2601(08)60317-2.
- Fu, G., Xu, F., Cameron, C. A., Heyman, G. D., & Lee, K. (2007). Cross-cultural differences in children's choices, categorizations, and evaluations of truths and lies. *Developmental Psychology, 43*, 278–293. doi: 10.1037/0012-1649.43.2.278.

- Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday life*. New York: Oxford University Press.
- Gelman, S. A., & Davidson, N. (2013). Conceptual influences on category-based induction. *Cognitive Psychology*, *66*, 327-353. doi: 10.1016/j.cogpsych.2013.02.001
- Gelman, S. A., & Heyman, G. D. (1999). Carrot-eaters and creature-believers: The effects of lexicalization on children's inferences about social categories. *Psychological Science*, *10*, 489-493. doi:10.1111/1467-9280.00194
- Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition*, *23*, 183-209.
- Gelman, S. A., Ware, E. A., & Kleinberg, F. (2010). Effects of generic language on category content and structure. *Cognitive Psychology*, *61*, 273-301.
doi:10.1016/j.cogpsych.2010.06.001
- Graham, S. A., Keates, J., Vukatana, E., & Khu, M. (2013). Distinct labels attenuate 15-month-olds' attention to shape in an inductive inference task. *Frontiers in Psychology*, *3*, 1-8.
doi: 10.3389/fpsyg.2012.00586
- Graham, S. A., & Kilbreath, C. S. (2007). It's a sign of the kind: Gestures and words guide infants' inductive inferences. *Developmental Psychology*, *43*, 1111-1123. doi:
10.1037/0012-1649.43.5.1111
- Graham, S. A., Nayer, S. L., & Gelman, S. A. (2011). Two-year-olds use the generic/nongeneric distinction to guide their inferences about novel kinds. *Child Development*, *82*, 493-507.
doi: 10.1111/j.1467-8624.2010.01572.x

- Göckeritz, S., Schmidt, M. F. H., & Tomasello, M. (2014). Young children's creation and transmission of social norms. *Cognitive Development, 30*, 81–95. doi: 10.1016/j.cogdev.2014.01.003
- Hammond, M. D., & Cimpian, A. (2017). Investigating the cognitive structure of stereotypes: Generic beliefs about groups predict social judgments better than statistical beliefs. *Journal of Experimental Psychology: General*. Manuscript in press.
- Hardecker, S., Schmidt, M. F. H., Roden, M., & Tomasello, M. (2016). Young children's behavioral and emotional responses to different social norm violations. *Journal of Experimental Child Psychology, 150*, 364–379. doi: 10.1016/j.jecp.2016.06.012
- Hart, D., & Edelstein, W. (1992). Self-understanding development in cross-cultural perspective. In T. M. Brinthaupt & R. P. Lipka (Eds.), *The self: Definitional and methodological issues* (pp. 291–322). New York: State University of New York Press.
- Haun, D. B. M., & Tomasello, M. (2011). Conformity to peer pressure in preschool children. *Child Development, 82*, 1759–1767. doi:10.1111/j.1467-8624.2011.01666.x
- Heilman, M. E. (2012). Gender stereotypes and workplace bias. *Research in Organizational Behavior, 32*, 113–135. doi:10.1016/j.riob.2012.11.003
- Heiphetz, L., Spelke, E. S., Harris, P. L., & Banaji, M. R. (2013). The development of reasoning about beliefs: Fact, preference, and ideology. *Journal of Experimental Social Psychology, 49*, 559–565. doi: 10.1016/j.jesp.2012.09.005
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33*, 61–135.

- Ho, A. K., Roberts, S. O., & Gelman, S. A. (2015). Essentialism and racial bias jointly contribute to the categorization of multiracial individuals. *Psychological Science, 26*, 1–7.
doi:10.1177/0956797615596436
- Horner, V., & Whiten, A. (2005). Causal knowledge and imitation/emulation switching in chimpanzees (*Pan troglodytes*) and children (*Homo sapiens*). *Animal Cognition, 8*, 164–181. doi: 10.1007/s10071-004-0239-6
- Hume, D. (2000). *A treatise of human nature* (D. F. Norton & M. J. Norton, Eds.). New York, NY: Oxford University Press. (Original work published 1740)
- Huynh, H., & Feldt, L. S. (1976). Estimation of the Box correction for degrees of freedom from sample data in randomized block and split-plot designs. *Journal of Educational Statistics, 1*, 69–82. doi:10.3102/10769986001001069.
- Josephs, M., Kushnir, T., Gräfenhain, M., & Rakoczy, H. (2016). Children protest moral and conventional norms when they believe actions are freely chosen. *Journal of Experimental Child Psychology, 141*, 247-255. doi: 10.1016/j.jecp.2015.08.002
- Jussim, L. (2012). *Social perception and social reality: Why accuracy dominates bias and self-fulfilling prophecy*. New York: Oxford University Press.
- Kalish, C. W. (2012). Generalizing norms and preferences within social categories and individuals. *Developmental Psychology, 48*, 1133–1143. doi:10.1037/a0026344
- Keil, F. C. (1989). *Concepts, kinds, and cognitive development*. Cambridge, MA: Bradford Book/MIT Press.
- Kenward, B. (2012). Over-imitating preschoolers believe unnecessary actions are normative and enforce their performance by a third party. *Journal of Experimental Child Psychology, 112*, 195–207. doi:10.1016/j.jecp.2012.02.006

- Kurzban, R., Tooby, J., & Cosmides, L. (2001). Can race be erased? Coalitional computation and social categorization. *Proceedings of the National Academy of Sciences*, 98, 15387–92. doi: 10.1073/pnas.251541498
- Legare, C. H., & Harris, P. L. (2016). The ontogeny of cultural learning. *Child Development*, 87, 633–642. doi: 10.1111/cdev.12542
- Levy, G. D., Taylor, M. G., & Gelman, S. A. (1995). Traditional and evaluative aspects of flexibility in gender roles, social conventions, moral rules, and physical laws. *Child Development*, 66, 515–531. doi:10.1111/j.1467-8624.1995.tb00887.x
- Lickel, B., Hamilton, D. L., Wierzchowska, G., Lewis, A., Sherman, S. J., & Uhles, A. N. (2000). Varieties of groups and the perception of group entitativity. *Journal of Personality and Social Psychology*, 78, 223–246. doi: 10.1037/0022-3514.78.2.223
- Liben, L. S., Bigler, R. S., & Krogh, H. R. (2001). 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*, 79, 346–63. doi: 10.1006/jecp.2000.2611
- Liben, L. S., Bigler, R. S., & Krogh, H. R. (2001). 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*, 79, 346–363. doi:10.1006/jecp.2000.2611
- Liben, L. S., Bigler, R. S., & Krogh, H. R. (2002). Language at work: Children’s gendered interpretations of occupational titles. *Child Development*, 73, 810–828.
- Lickel, B., Hamilton, D. L., Wierzchowska, G., Lewis, A., Sherman, S. J., & Uhles, A. N. (2000). Varieties of groups and the perception of group entitativity. *Journal of Personality and Social Psychology*, 78, 223–246. doi:10.1037/0022-3514.78.2.223

- Little, E. E., Carver, L. J., & Legare, C. H. (2016). Cultural variation in triadic infant-caregiver object exploration. *Child Development, 87*, 1130-1145. doi: 10.1111/cdev.12513.
- Lyons, D. E., Damrosch, D. H., Lin, J. K., Macris, D. M., & Keil, F. C. (2011). The scope and limits of overimitation in the transmission of artefact culture. *Philosophical Transactions of the Royal Society, 1158–1167*. doi: 10.1098/rstb.2010.0335
- Macnamara, J. (1982). *Names for things*. Cambridge, MA: MIT Press.
- Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review, 98*, 224–253. doi:10.1037/0033-295X.98.2.224
- McAuliffe, K., & Dunham, Y. (2016). Group bias in cooperative norm enforcement. *Philosophical Transactions of the Royal Society B: Biological Sciences, 371*, 1–9. doi:10.1098/rstb.2015.0073
- McGuire, L., Rutland, A., & Nesdale, D. (2015). Peer group norms and accountability moderate the effect of school norms on children’s intergroup attitudes. *Child Development, 86*, 1290–1297. doi: 10.1111/cdev.12388
- Miller, P. J., Wiley, A. R., Fung, A., Liang, C. (1997). Personal storytelling as a medium of socialization in Chinese and American families. *Child Development, 68*, 557-568. doi: 10.1111/1467-8624.ep9709050634
- Modyanova, N., Agoos, C., Kenney, A., Echelbarger, M., Holt, A., & Wexler, K. (2010). Young children’s interpretations of modal verbs. In J. Costa, A. Castro, M. Lobo & F. Pratas (Eds.). *Language Acquisition and Development: Proceedings of GALA 2009*. Cambridge Scholars Press. pp. 301–3012.
- Mulvey, K. L. (2016). Evaluations of moral and conventional intergroup transgressions. *British Journal of Developmental Psychology, 34*, 489-501. doi:10.1111/bjdp.12145

- Mulvey, K. L., Hitti, A., Rutland, A., Abrams, D., & Killen, M. (2014). When do children dislike ingroup members? Resource allocation from individual and group perspectives. *Journal of Social Issues, 70*, 29–46. doi:10.1111/josi.12045
- Nielsen, M., & Haun, D. (2016). Why developmental psychology is incomplete without comparative and cross-cultural perspectives. *Philosophical Transactions of the Royal Society B: Biological Sciences, 371*, 1–7. doi:10.1098/rstb.2015.0069
- Nielsen, M., & Mushin, I. (2014). Where culture takes hold: “Overimitation” and its flexible deployment in Western, Aboriginal, and Bushmen Children. *Child Development, 85*, 2169–2184. doi: 10.1111/cdev.12265
- Nielsen, M., & Tomaselli, K. (2010). Overimitation in Kalahari Bushman children and the origins of human cultural cognition. *Psychological Science, 21*, 729–736. doi: 10.1177/0956797610368808
- Olson, K. R., & Dweck, C. S. (2008). A blueprint for social cognitive development. *Perspectives on Psychological Science, 3*, 193–202.
- Oyserman, D., Coon, H. M., & Kemmelmeier, M. (2002). Rethinking individualism and collectivism: Evaluation of theoretical assumptions and meta-analyses. *Psychological Bulletin, 128*, 3–72. doi: 10.1037//0033-2909.128.1.3
- Papafragou, A. (1997). Modality in language development: A reconsideration of the evidence. *University College London Working Papers in Linguistics, 9*, 1–31. Retrieved on April 25th, 2016 from: <http://papafragou.psych.udel.edu/papers/uclwp97.pdf>
- Piatraszewski, D., Cosmides, L., & Tooby, J. (2014). The content of our cooperation, not the color of our skin: An alliance detection system regulates categorization by coalition and race, but not sex. *PLOS ONE, 9*, 1-19. doi: 10.1371/journal.pone.0088534

- Pingali, P. (2006). Westernization of Asian diets and the transformation of food systems: Implications for research and policy. *Food Policy*, *32*, 281–298. doi: 10.1016/j.foodpol.2006.08.00
- Plötner, M., Over, H., Carpenter, M., & Tomasello, M. (2016). What is a group? Young children's perceptions of different types of groups and group entitativity. *PLOS ONE*, *11*, 1–14. doi:10.1371/journal.pone.0152001
- Prasada, S. (2000) Acquiring generic knowledge. *Trends in Cognitive Sciences*, *4*, 66–72. doi: 10.1016/S1364-6613(99)01429-1
- Quintana, S. M. (1998). Children's developmental understanding of ethnicity and race. *Applied and Preventive Psychology*, *7*, 27–45. doi:10.1016/S0962-1849(98)80020-6
- Rakoczy, H., Hamann, K., Warneken, F., & Tomasello, M. (2010). Bigger knows better: Young children selectively learn rule games from adults rather than from peers. *British Journal of Developmental Psychology*, *28*, 785–798. doi: 10.1348/026151009X479178
- Rakoczy, H., & Schmidt, M. F. H. (2013). The early ontogeny of social norms. *Child Development Perspectives*, *7*, 17–21. doi:10.1111/cdep.12010
- Rakoczy, H., Warneken, F., & Tomasello, M. (2008). The sources of normativity: Young children's awareness of the normative structure of games. *Developmental Psychology*, *44*, 875-881. doi: 10.1037/0012-1649.44.3.875
- Rhodes, M. (2012). Naive theories of social groups. *Child Development*, *83*, 1900–1916. doi:10.1111/j.1467-8624.2012.01835.x
- Rhodes, M. (2014). Children's explanations as a window into their intuitive theories of the social world. *Cognitive Science*, *38*, 1687–1697. doi:10.1111/cogs.12129

- Rhodes, M., & Chalik, L. (2013). Social categories as markers of intrinsic interpersonal obligations. *Psychological Science, 24*, 999–1006. doi:10.1177/09567976124466267
- 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. doi:10.1016/j.cogpsych.2009.05.001
- Rhodes, M., Leslie, S.-J., & Tworek, C. M. (2012). Cultural transmission of social essentialism. *Proceedings of the National Academy of Sciences of the United States of America, 109*, 13526–13531. doi:10.1073/pnas.1208951109
- Riggs, A. E., & Young, A. G. (2016). Developmental changes in children’s normative reasoning across learning contexts and collaborative roles. *Developmental Psychology, 52*, 1236–1246. doi: 10.1037/dev0000119
- Roberts, S. O., & Gelman, S. A. (2015). Do children see in Black and White? Children’s and adults’ categorizations of multiracial individuals. *Child Development, 86*, 1830–1847. doi:10.1111/cdev.12410
- 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. doi: 10.1037/dev000132
- Roberts, S. O., & Gelman, S. A. (2017a). Multiracial children’s and adults’ categorizations of multiracial individuals. *Journal of Cognition and Development, 18*, 1-15. doi: 10.1080/15248372.2015.1086772
- Roberts, S. O., & Gelman, S. A. (2017b). Now you see race, now you don’t: Verbal cues influence children’s racial stability judgments. *Cognitive Development, 43*, 129-141. doi: 10.1016/j.cogdev.2017.03.003

- Roberts, S. O., Gelman, S. A., & Ho, A. K.. (2017a). So it is, so it shall be: Group regularities license children's prescriptive judgments. *Cognitive Science*, *41*, 576-600. doi: 10.1111/cogs.12443
- Roberts, S. O., Guo, C., Ho, A. K., & Gelman, S. A. (2017c). Children's descriptive-to-prescriptive replicates (and varies) cross-culturally. *Journal of Experimental Child Psychology*. Advanced online publication.
- Roberts, S. O., Ho, A. K., & Gelman, S. A. (2017b). Group presence, category labels, and generic statements influence children to treat descriptive group regularities as prescriptive. *Journal of Experimental Child Psychology*, *158*, 19-31. doi: 10.1016/j.jecp.2016.11.013
- Rudman, L. A., Moss-Racusin, C. A., Glick, P., & Phelan, J. E. (2012). Reactions to vanguards: Advances in backlash theory. In Devine , P. G. , Plant & E. A (Eds.), *Advances in Experimental Social Psychology*, *45*, 167–227.
- Rutland, A., Hitti, A., Mulvey, K. L., Abrams, D., & Killen, M. (2015). When does the in-group like the outgroup? Bias among children as a function of group norms. *Psychological Science*, *26*, 834–842. doi:10.1177/0956797615572758
- Samland, J., Josephs, M., Waldmann, M. R., & Rakoczy, H. (2016). The role of prescriptive norms and knowledge in children's and adults' causal selection. *Journal of Experimental Psychology: General*, *145*, 125–130. doi.org/10.1037/xge0000138
- Schillaci, R. S., & Kelemen, D. (2014). Children's conformity when acquiring novel conventions: The case of artifacts. *Journal of Cognition and Development*, *15*, 569–583. doi:10.1080/15248372.2013.784973

- Schmidt, M. F. H., Butler, L. P., Heinz, J., & Tomasello, M. (2016). Young children see a single action and infer a social norm. Promiscuous normativity in 3-year-olds. *Psychological Science*, 27, 1360-1370. doi: 10.1177/0956797616661182
- Schmidt, M. F. H., Hardecker, S., & Tomasello, M. (2016). Preschoolers understand the normativity of cooperatively structured competition. *Journal of Experimental Child Psychology*, 143, 34–47. doi: 10.1016/j.jecp.2015.10.014
- Schmidt, M. F. H., Rakoczy, H., & Tomasello, M. (2011). Young children attribute normativity to novel actions without pedagogy or normative language. *Developmental Science*, 14, 530–539. doi: 10.1111/j.1467-7687.2010.01000.x
- Sherif, M., & Sherif, C. W. (1969). *Social Psychology*. New York: Harper & Row.
- Shutts, K., Pemberton Roben, C. K., & Spelke, E. S. (2013). Children’s use of social categories in thinking about people and social relationships. *Journal of Cognition and Development*, 14, 35–62. doi:10.1080/15248372.2011.638686
- Smetana, J. G. (1981). Preschool children’s conceptions of moral and social rules. *Child Development*, 52, 1333-1336. doi: 10.2307/1129527
- Smetana, J. G., Jambon, M., Conry-Murray, C., & Sturge-Apple, M. L. (2012). Reciprocal associations between young children’s developing moral judgments and theory of mind. *Developmental Psychology*, 48, 1144–1155. doi:10.1037/a0025891
- Spielman, D. A. (2000). Young children, minimal groups, and dichotomous categorization. *Personality & Social Psychology Bulletin*, 26, 1433–1441.. doi: 10.1177/0146167200263010

- Strachan, J., Samuel, J., & Takaro, M. (2007). Ni Vanuatu women graduates: What happens when they go home? *Development in Practice, 17*, 147–153. doi: 10.1080/09614520601092055
- Tajfel, H., & Turner, J.C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33–47). Monterey, CA: Brooks/Cole.
- Talhelm, T., Zhang, X., Oishi, S., Shimin, C., Duan, D., Lan, X., & Kitayama, S. (2014). Large-scale psychological rice versus wheat agriculture. *Science, 344*, 603–608. doi: 10.1126/science.1246850
- 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. doi: 10.1111/j.1467-8624.2009.01272.x
- Tworek, C. M., & Cimpian, A. (2016). Why do people tend to infer ought from is? The role of biases in explanation. *Psychological Science, 27*, 1109–1122. doi: 10.1177/0956797616650875
- Vaish, A., Hermann, E., Markmann, C., & Tomasello, M. (2016). Preschoolers value those who sanction non-cooperators. *Cognition, 153*, 43-51. doi: 10.1016/j.cognition/2016.04.011
- Walton, G. M., & Banaji, M. R. (2004). Being what you say: The effect of linguistic labels on attitudes. *Social Cognition, 22*, 193–213. doi: 10.1521/soco.22.2.193.35463
- Wang, Q. (2016). Why should we all be cultural psychologists? Lessons from the study of social cognition. *Perspectives on Psychological Science, 11*, 583-596. doi: 10.1177/1745691616645552

- Wang, Q., Leichtman, M. D., & Davies, K. (2000). Sharing memories and telling stories: American and Chinese mothers and their 3-year-olds. *Memory*, 8, 159–177. doi: 10.1080/096582100387588
- Waxman, S. R., & Markow, D. B. (1995). Words as invitations to form categories: evidence from 12-month old infants. *Cognitive Psychology*, 29, 257-302. doi: 10.1006/cogp.1995.1016
- Waxman, S. R. (2010). Names will never hurt me? Naming and the development of racial and gender categories in preschool-aged children. *European Journal of Social Psychology*, 40, 593–610. doi:10.1002/ejsp
- Wellman, H. W. (2014). *Making minds: How theory of mind develops*. New York: Oxford University Press.
- Wu, M. (2009). The transformation of the educational semantic within a changing society: A study of the Westernization of modern Chinese education. *Journal of Historical Sociology*, 22, 528-552. doi: 10.1111/j.1467-6443.2009.01368.
- Zeng, R., & Greenfield, P. M. (2015). Cultural evolution over the last 40 years in China: Using the Google Ngram Viewer to study implications of social and political change for cultural values. *International Journal of Psychology*, 50, 47-55. doi: 10.1002/ijop.12125