Title

How Everyday Language Shapes the Development of Stereotypes

Author

Baharloo, Roya

Publication Date

2023

Peer reviewed|Thesis/dissertation

How Everyday Language Shapes the Development of Stereotypes

By

Roya Baharloo

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Psychology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Mahesh Srinivasan, Chair
Professor Arianne Eason
Professor Jan Engelmann
Professor Lin Bian

Fall 2023

How Everyday Language Shapes the Development of Stereotypes

by

Roya Baharloo

Abstract

How Everyday Language Shapes the Development of Stereotypes

by

Roya Baharloo

Doctor of Philosophy in Psychology

University of California, Berkeley

Professor Mahesh Srinivasan, Chair

Stereotypes are powerful. They not only influence our perceptions and interactions but also shape individual identities and societal structures. Crucially, stereotypes emerge early, shaping children's understanding of the world and their place in it. The early emergence is particularly concerning as it underpins the perpetuation of social inequities, such as racial and gender disparities. This dissertation explores the development of stereotypes, focusing on the linguistic factors that contribute to their formation. By examining how language–both in terms of what is said about and to social groups–influences children's stereotypes, this research highlights the subtle yet powerful ways that everyday language can shape social cognition.

Chapter 1 examines the emergence of racial stereotypes in early childhood, revealing that by age 7, children start attributing competence and warmth differently across racial groups. This finding suggests these stereotypes are not due to mere in-group preferences, but rather, mirror broader societal beliefs. By shedding light on the developmental trajectory of stereotypes, this chapter serves as a foundation for further exploration into the mechanisms underlying this development.

Chapter 2 shifts the focus to pragmatic inferences, investigating how children and adults "read between the lines" of what a speaker says to form judgments about social groups. The chapter illustrates the powerful role of indirect language cues in stereotype formation and the nuanced ways language impacts social cognition.

Chapter 3 demonstrates that the manner of addressing social groups affects perceptions of their competence. The research reveals that both children and adults perceive individuals addressed with simplified language or information as less competent. This novel perspective expands our understanding of linguistic factors in stereotype formation, moving beyond the traditional focus on the content of speech about social groups, and highlights the significance of how speech is directed at them.

Collectively, this dissertation presents a unique perspective on the development of stereotypes, emphasizing the significant role of linguistic cues from childhood through adulthood. This work contributes to academic discourse and offers valuable insights for developing educational and policy interventions aimed at dismantling harmful stereotypes.

Table of Contents

Introduction

Stereotypes–broad generalizations about social groups–are prevalent in our society. Arising from our inherent need to simplify the overwhelmingly complex world around us, stereotypes serve as cognitive shortcuts for understanding our social environment. However, their impact extends far beyond mere mental categorizations, often leading to significant real-world consequences. This dissertation focuses on the development of stereotypes, revealing their early emergence in childhood and highlighting the powerful role of language in shaping these beliefs.

Stereotypes are Captured by Two Dimensions: Warmth and Competence

To understand the development of stereotypes, it is essential to examine their underlying structure. Stereotypes extend beyond simple judgments of "good" or "bad." Several theories, including the Stereotype Content Model, suggest that our social cognition is centered on two fundamental dimensions: warmth and competence (Fiske, 2018; Fiske et al., 2002). The warmth dimension refers to perceived intention–i.e., whether somebody has benign or harmful intentions–and is captured by traits such as friendliness, helpfulness, and sincerity. The competence dimension refers to perceived ability–i.e., how capable somebody is of acting on their intentions–and is captured by traits such as intelligence, creativity, and efficacy.

Consequences of Stereotypes

Stereotypes extend beyond cognitive heuristics about warmth and competence; they significantly influence prejudices and dictate real-world behaviors (Cuddy et al., 2007). Groups perceived as lacking in both dimensions often face societal contempt, whereas those seen as competent but not warm may evoke envy. Conversely, groups regarded as warm but not competent typically elicit pity. Importantly, stereotypes about competence can reinforce oppressive societal inequalities. For instance, groups that are stereotyped as having relatively lower intelligence, such as Black people and women, may feel discouraged from pursuing opportunities in fields that emphasize brilliance, such as math and philosophy (Leslie et al., 2015). In a different vein, groups perceived as highly competent, like Asian people, face their own set of challenges, such as undue academic pressure and heightened expectations (Kao, 2000). This can lead to negative effects on self-esteem and cognitive performance. Cheryan & Bodenhausen (2000) illustrate this, showing that Asian women's math test performance was undermined when they were subjected to high performance expectations, likely due to increased stress and a decreased ability to concentrate.

The Emergence of Stereotypes in Childhood

Understanding the development of stereotypes is crucial as childhood is when these harmful beliefs take root. By the age of 6, children have already internalized stereotypes, such as associating brilliance with males more than females. This not only leads young girls to perceive members of their own gender as less intellectually capable but also discourages them from participating in activities deemed for children who are "really, really smart" (Bian et al., 2017). As children's beliefs and behaviors are molded by societal stereotypes, they may unwittingly reinforce these biases and even exclude peers from "less competent" groups (Bian et al., 2018; Darley & Fazio, 1980). Targeting this stage, particularly for children who have not yet formed entrenched racial or gendered beliefs, presents a valuable opportunity to mitigate the long-term impact and perpetuation of harmful stereotypes.

The Mechanisms Underlying Stereotype Development

Considering that stereotypes emerge early in life and have a profound influence on our thoughts, actions, and society, a critical question arises: where do these stereotypes originate? The Developmental Intergroup Theory identifies several core cognitive processes that contribute to the formation of stereotypes and prejudice (Bigler & Liben, 2007). Initially, children discern personal attributes using cues from their environment, prioritizing classifications like gender over less socially salient attributes like handedness. Cues like perceptual differences, explicit labeling, implicit groupings, and group sizes aid in this determination. As children further engage with their environment, they begin to attach meaning to these noticeable groups. This process is influenced by both internal cognitive processes and external influences. Internally, children might link certain social categories with specific attributes or beliefs. For instance, due to essentialist thinking, they might reason that Black and White people have different blood types (S. Gelman, 2004). This kind of thinking can also create a natural preference for their own social group, leading to early forms of prejudice. Externally, children's beliefs are shaped by direct statements, such as "African Americans are hostile" as well as subtler cues, like noticing associations between particular social groups and certain characteristics.

Explicit and Implicit Attitudes

Building on this foundation, it is crucial to explore the external forces driving stereotype development. Central to this exploration are explicit and implicit attitudes. Explicit attitudes are conscious beliefs that individuals can articulate, often measured through self-report. While it might seem that overt statements like "African Americans are hostile" would be a primary, straightforward route for stereotype transmission, this isn't typically the case due to social norms that condemn blatant bias and prejudice (Dovidio, 2001). Conversely, implicit attitudes, which operate subconsciously, manifest as automatic evaluations triggered by the mere presence of a stimulus. Detected through techniques like response latency or physiological measures, these attitudes often persist even in societies that endorse explicit egalitarian values, subtly influencing behavior.

The Role of Language on Stereotype Transmission

Language has been recognized as a powerful tool for transmitting stereotypes. Even when the intent is neutral or positive, subtle linguistic cues can shape beliefs. For example, statements aimed at promoting gender equality, such as “Girls are as good as boys at math,” can inadvertently strengthen gender stereotypes due to the syntactical framing that sets one gender (in this case, boys) as the standard of ability (Chestnut et al., 2021; Chestnut & Markman, 2018). Generic language, such as "Boys are good at math," significantly influences children's knowledge about categories. These statements differ from specific language (e.g., "This boy is good at math") as they signal that a particular characteristic (e.g., good at math) applies to an entire kind (e.g., boys). Such generic statements are common in child-directed speech and can be understood by children as young as 2 or 3 years old (Cimpian, 2013; Cimpian & Markman, 2008, 2011; S. A. Gelman et al., 2008; S. A. Gelman & Raman, 2003). More recent research has demonstrated that children use generic language to learn about social groups, even if they are not explicitly mentioned. For example, after hearing that “Gorps are good at baking pizza,” children are more likely to think that a new Gorp is good at baking pizzas but are also more likely to infer that other members of another group, Zarpies, might not be (Moty & Rhodes, 2021).

Language Variation in Response to Audience Characteristics

Stereotypes don't just shape our perceptions; they also impact how we communicate. Effective communication often requires adjusting one's language to match the listener's perceived abilities or identity (Bell, 1984; Clark, 1998). Such adjustments are often made subconsciously and can be deeply influenced by societal stereotypes. Instances of accommodation, where speakers modify their speech to manage social distances, can lead to stereotypical implications. For instance, a younger person who oversimplifies their speech when addressing an older individual may unintentionally convey perceptions of the elderly as less competent. This "patronizing speech," even if unintentional, mirrors harmful stereotypes and biases (Giles & Ogay, 2007). The subtle ways in which language adjustments can reveal underlying stereotypes are exemplified in the work of Dupree & Fiske (2019). They found that White liberals, aiming to affiliate with Black people, unconsciously used fewer competence-related words, reflecting negative competence stereotypes. These phenomena underscore the significant influence of stereotypes on how we use language.

Research Focus

This dissertation examines how language indirectly shapes the development of stereotypes, with a focus on the role of linguistic adjustments. It specifically investigates how observers perceive a listener's competence when witnessing such adjustments in language, offering insights into the perpetuation of stereotypes through everyday communication.

Chapter 1: The Development of Racial Stereotypes about Warmth and Competence

Chapter 1 examines the emergence of racial stereotypes in early childhood, revealing that by age 7, children start attributing competence and warmth differently across racial groups. This finding suggests these stereotypes are not due to mere in-group preferences, but rather, mirror broader societal beliefs. By shedding light on the developmental trajectory of stereotypes, this chapter serves as a foundation for further exploration into the mechanisms underlying this development.

Chapter 2: Children's Use of Pragmatic Inference to Learn about the Social World
Chapter 2 shifts the focus to pragmatic inferences, investigating how children and adults "read between the lines" of what a speaker says to form judgments about social groups. The chapter illustrates the powerful role of indirect language cues in stereotype formation and the nuanced ways language impacts social cognition.

Chapter 3: Children Infer the Relative Competence of Individuals by Observing How They Are Addressed
Chapter 3 demonstrates that the manner of addressing social groups affects perceptions of their competence. The research reveals that both children and adults perceive individuals addressed with simplified language or information as less competent. This novel perspective expands our understanding of linguistic factors in stereotype formation, moving beyond the traditional focus on the content of speech about social groups, and highlights the significance of how speech is directed at them.

Chapter 1: The Development of Racial Stereotypes About Warmth and Competence

Abstract

Stereotypes about social groups are captured by two dimensions: warmth and competence (the Stereotype Content Model). In the U.S., Black people are stereotyped as less competent and less warm than White people, whereas Asian people are stereotyped as more competent but less warm than White people. These racial stereotypes influence how group members fare within society. The present research tested whether children subject to the stereotype content model to construct their racial beliefs. In Experiment 1, children ages 5 to 7 (*N* = 72, 36 girls; 47% White) saw pictures depicting White people and Black people and guessed who was “really nice” or “really smart” (child-friendly terms for warmth and competence). Five and 6-year-olds tended to select White people as being warm and competent, whereas 7-year-olds attributed competence, but not warmth, to White people. In Experiment 2, children ages 5 to 7 (*N* = 72, 36 girls; 55% White) saw pictures depicting White people and Asian people. Although 5- and 6-year-olds were equally likely to attribute warmth and competence to White people and Asian people, 7-year-olds began to associate competence with Asian people and warmth with White people. These results suggest that children’s racial stereotypes are not merely based on a general preference for racial ingroup members or White people. Rather, children apply the Stereotype Content Model to differentially attribute warmth and competence to different racial groups. These findings add to our knowledge about stereotype development and have broad implications for understanding the roots of the current racial disparities.

Introduction

The U.S. has seen a decline in overt expressions of racial prejudice over the past few decades, partially due to changes in social norms and legislation (Dovidio, 2001). However, subtle forms of racial attitudes—such as stereotypes—persist. Research with adults has demonstrated that common stereotypes about social groups are captured by two dimensions: warmth and competence. In the U.S., where White people are generally seen as warm and competent, Black people are perceived as relatively less warm and less competent, and Asian people are perceived as relatively less warm but more competent (Fiske et al., 2002). These racial stereotypes affect how members of different racial groups fare within society, for example, by influencing intergroup wellbeing (e.g., Cuddy et al., 2007; Fiske et al., 2002) and career and educational outcomes (e.g., Leslie et al., 2015; Meyer et al., 2015). Yet, less is known about the development of these stereotypes, posing challenges to alleviate these harmful beliefs from their roots. In the present research, we examined when children endorse stereotypical beliefs associating warmth and competence with different races.

Research with adults has demonstrated that when judging others, adults do not simply view individuals as "good" or "bad." Rather, as suggested by the Stereotype Content Model (SCM), common stereotypes rely on two important dimensions: warmth and competence (Fiske, 2018; Fiske et al., 2007). The warmth dimension refers to perceived intention–i.e., whether somebody has benign or harmful intentions–and is captured by traits such as friendliness, helpfulness, and sincerity. The competence dimension refers to perceived ability–i.e., how capable somebody is of acting on their intentions–and is captured by traits such as intelligence, creativity, and efficacy. Similar to other social categories, racial groups are differentially stereotyped along these dimensions. For example, compared to White people, who are generally seen as competent and warm, Black people are generally perceived as relatively less competent and less warm, and Asian people are perceived as relatively more competent and less warm (Bergsieker et al., 2012; Fiske et al., 2002; Park et al., 2015). Importantly, these stereotypes bring in myriads of consequences. Groups who are stereotyped as low in warmth and competence—such as Black people—are likely to be targets of contemptuous prejudice, which involves feelings of anger, contempt, and disgust (Fiske et al., 2002). In contrast, groups who are stereotyped as low in warmth and high in competence—such as Asian people—are likely to be targets of envious prejudice, which involves feelings of envy and jealousy (Fiske, 2015; Fiske et al., 2002; Lee & Fiske, 2006). The distinct prejudices are translated into actual behaviors towards different groups (Fiske et al., 2007).

Recently, stereotypes about competence have been found to be at least partially responsible for the group inequality in the current society (e.g., Leslie et al., 2015; Meyer et al., 2015). For example, groups that are stereotyped as having relatively low intelligence, such as Black people and women, may be discouraged from pursuing opportunities in fields that emphasize brilliance such as math and philosophy (Leslie et al., 2015). Stereotypes about high competence are similarly harmful. Groups that are stereotyped as being intelligent, such as Asian people, may feel extra pressure to perform well academically and find it difficult to feel content with themselves (Kao, 2000). Further, this added pressure may lead to impaired intellectual performance. For example, Cheryan & Bodenhausen (2000) demonstrated that highlighting others' expectations of Asians' excellent performance undermined Asian women's performance on a math test due to a decreased ability to concentrate.

When do children develop these pernicious stereotypes linking warmth and competence with some racial groups, but not others? Exploring this question would contribute to our understanding of the developmental roots of the racial disparity and inform educators and parents of the precise timeline to implement interventions. Moreover, it helps to advance the current theories of stereotyping in development. Existing theories uncover the acquiring mechanisms underlying the development of stereotypes. For example, Developmental Intergroup Theory (DIT; Bigler & Liben, 2007) proposes several cognitive processes that lead children to become aware of race and develop conceptual categories based on race. However, there is limited work investigating the content of early racial stereotypes and the timeline that these stereotypes become entrenched.

Children are sensitive to race at an early age. Three-month-old infants can perceptually discriminate racial groups and prefer to look at faces that are familiar in their environment (Bar-Haim et al., 2006; Kelly et al., 2009). By the time children are 3-4 years of age, they can sort people by race and favor their racial ingroup members over members of racial outgroups (Aboud, 2003; Castelli et al., 2007; Nesdale, 2001). By the time White children are in kindergarten, they endorse implicit and explicit racial ingroup preferences (Baron & Banaji, 2006; Cameron et al., 2001; Nesdale, 2001; Shutts, 2015). Children apply positive and negative stereotypes to racial outgroup members by the time they are 6 years of age (Pauker et al., 2010). More relevant to racial beliefs about intellectual skills, the findings are more mixed. One study suggests that children in fifth grade begin to attribute high math abilities to Asian people (Cvencek et al., 2015), whereas a recent study finds that 4- to 8-year-olds sometimes rated Asians as less smart and sometimes as equally smart relative to White people (Sierksma et al., 2022). Open questions remain whether and when children endorse common racial stereotypes about warmth and competence.

Investigating these questions also provides useful insights on whether children hold a global or a more fine-grained view of attributes such as warmth and competence. Much work has suggested that young children have difficulties differentiating evaluative traits (Heyman et al., 2003; Ruble & Dweck, 1995). For example, first graders tended to refer to prosocial behaviors when defining the term “smart” (Yussen & Kane, 1985). Similarly, children seem to conflate warmth and competence in their perceptions of social groups until late childhood, sometime after 10 years of age (Roussos & Dunham, 2016). However, these results may underestimate children’s ability to reason about these traits (e.g., by asking children to predict one’s resource allocations as an indicator of their warmth judgments). Recent studies that presented children with exemplars of warmth and competence, as well as more direct questions, provide evidence for an early-emerging, more fine-grained understanding of these characteristics (Bian et al., 2017; Cimpian et al., 2007). For example, although children at the age six attribute being smart to White men/boys, they tend to associate niceness with White women/girls (Bian et al., 2017; Shu et al., 2022). These recent studies make it possible to probe warmth vs competence as dimensions of racial attitudes in children.

In addition, we explored how children’s own race influences their racial stereotypes. Previous theories such as Social Identity Theory (SIT; Tajfel, 1970) and the Social Reasoning Developmental perspective (SRD; Rutland et al., 2010) suggest that children’s group identity plays a role in their social reasoning. For example, according to SIT, people have a general preference for their ingroups over outgroups to maintain a positive social identity. As noted earlier, White children from early on prefer members of their own race over members of other races (Baron & Banaji, 2006; Dunham et al., 2006; Kinzler et al., 2009; McGlothlin et al., 2005; Raabe & Beelmann, 2011). However, children of other racial and ethnic backgrounds do not demonstrate this ingroup preference, presumably because they hold favorable attitudes towards White people, a group that is linked with high social status (e.g., Newheiser et al., 2014; Shutts et al., 2011). The current research investigates children’s racial beliefs about warmth and competence, two highly valuable attributes in the U.S. If children go beyond a general preference for their ingroup members or White people, and are instead subject to SCM to shape their beliefs, we expect them to hold similar racial beliefs about warmth and competence regardless of their own racial identity. Finding that children evaluate non-White people as equally (or more) warm and competent as White people in some contexts would also provide evidence for children's sensitivity to SCM. Given the importance of this open question, despite our relatively small sample size, we consider children's own racial identity investigating the development of racial stereotypes.

In the current research, we focus on specific areas of warmth and competence: niceness and smartness, respectively–important aspects of warmth and competence that are well-understood by children (Bian et al., 2017). Specifically, we examined when children endorse the stereotypical beliefs that warmth and competence are qualities associated with different races. Given previous research suggesting that children in early elementary school years develop robust sensitivity to race (Dunham et al., 2006; Raabe & Beelmann, 2011) and endorse stereotypes about intellectual abilities (Bian et al., 2017; Cvencek et al., 2015), we tested children ages 5 to 7 years. In Experiment 1, we presented children with White individuals and Black individuals to explore their racial stereotypes about warmth and competence. In Experiment 2, we test whether children associate lower warmth and higher competence with Asian individuals compared to White individuals.

Data collection started in December 2018 and ended in March 2020. These experiments were not pre-registered. All study materials, data, analyses, and supplemental materials can be accessed on Open Science Framework: https://osf.io/hrm2t/.

Experiment 1: Perceptions of White and Black Individuals

Method

Power Analysis

We conducted a priori power analysis (G*Power 3.1; Faul et al., 2007) for a regression model with three predictors (i.e., participant age, target trait, and their interaction). We predicted an interaction between age (continuous) and trait (categorical). Based on past studies investigating children's gender stereotypes about competence that found an interaction between age (continuous) and gender (categorical) (e.g., Shu et al., 2022), we specified an effect size of f^2 = 0.17 with alpha set at .05. The analysis suggested that the minimum number of participants needed to provide 80% power to detect significant predictors was 69. We included 72 children in order to properly counterbalance all aspects of the study (i.e., gender, order of questions, etc.).

Participants

Our sample consisted of 72 children (24 per age group; 36 girls and 36 boys) aged 5 (M = 5.65 years), 6 (M = 6.49 years), and 7 (M = 7.50 years). Children were recruited from a database of families who were interested in research opportunities, as well as from local elementary schools and a museum. All children were tested at or near a major university in the Western United States, either in the laboratory (n = 28), local elementary schools (n = 27) or a local children's museum (n = 17) and received a small toy as compensation. An additional 10 children were tested but excluded from the final analysis because they did not pass the screener task (see below). Caregivers provided racial demographic information for 51 children. Of these children, 47% were White, 25% were multiracial (i.e., selected more than one race), 22% were Asian, and 6% were Black or African American.

Procedure and Measures

The procedure consisted of a screener task and two stereotype tasks adapted from Bian et al. (2017). In the screener task, the experimenter presented 12 questions to gauge children's understanding of niceness and smartness. Next, children were presented with two stereotype tasks assessing their racial stereotypes about niceness and smartness. In the stereotype tasks, children were shown pictures depicting White and Black faces, and were asked to indicate which one is "really really smart" or "really really nice." We adopted relative measures for several important reasons: First, stereotypes about racial groups are often relational in nature; Second, these measures are relatively implicit to camouflage our purpose of assessing children's racial stereotypes; Third, children at this age tend to be overly positive when making trait attributions (e.g., Boseovski, 2010), and using relative measures allow children to reveal their reasoning about groups (e.g., Ahl & Dunham, 2019).

Screener Task

The study began with a set of 12 questions intended to evaluate whether children understand the meaning of our target traits, "nice" (6 questions) and "smart" (6 questions). The questions were presented as separate blocks—one for nice, one for smart. The question order and block order were counterbalanced across participants. For each screener question, the experimenter described a child who was unknown to the participant while looking at a picture that was hidden from the child (e.g., "The child in this picture can always answer even the hardest questions from the teacher.") and then asked participants to indicate whether the target trait described the child (e.g., "Is this child smart, not smart, or are you not sure?"). Participants were not shown the pictures of the target children so that their responses to the following stereotype tasks would not be influenced by the children's racial identities. After each response, the experimenter either confirmed that they were correct, or corrected them if they responded incorrectly. We included data for children who answered at least 4 out of 6 screener questions correctly for each trait.

Stereotype Tasks

After the screener task, the experimenter administered 2 Stereotype Tasks that have been used in past work. We incorporated both tasks in order to provide a comprehensive assessment of children's racial beliefs. The order of the two tasks was counterbalanced across participants.

Stereotype Task 1. In Stereotype Task 1, the experimenter told 2 stories in counterbalanced order. One story was about somebody "really, really smart," and the other was about somebody "really, really nice." After each story, the experimenter presented participants with 4 pictures—2 of White individuals and 2 of Black individuals, all matched to the participants' gender. We selected pictures from the Chicago Face Database Version 2.0.3 (Ma et al., 2015), which contains high-resolution, standardized photographs of racially diverse men and women. The database contains norming data from 1,087 raters (a convenience sample from diverse racial backgrounds, with an average age of 26.75 years). In our sample of photos, each had ratings from an average of 44 raters. We selected pictures of targets who were rated as clearly White or Black, male or female, and we matched average ratings of perceived attractiveness and happiness across race. The experimenter presented the pictures to children in a line, alternating by race, then asked participants to guess which one of the 4 people they thought was the person in the story. For each trial, children received a score of 1 if they selected a White person and a 0 if they selected a Black person.

Stereotype Task 2. In Stereotype Task 2, participants were shown 6 pairs of pictures across 6 trials. The first 2 trials, which consisted of same-race picture pairs, served as practice trials. For the 4 test trials (Trials 3-6), each pair of pictures consisted of one White individual and one Black individual. In 2 of the trials, children were told that one of the two people was "really, really smart," and in the other 2 trials they were told that one of the two was "really, really nice". Children were asked to guess which of the two people had the relevant trait. The picture selection process was identical to the one described above for Stereotype Task 1. In each test trial, children received a score of 1 if they selected a White person and a 0 if they selected a Black person.

Analytic Strategy

Our primary analysis focused on whether children differentially associated smartness and niceness with White individuals and Black individuals, and if so, when this association developed. For each participant, we calculated a stereotype score, which was the proportion of trials (out of 3) that they selected a White person for each trait (nice or smart). We submitted the stereotype scores into a linear regression model in R (Version 3.6.1; R Core Team, 2019), including children's age (continuous variable), target trait (nice or smart), their interaction as fixed factors, and a random intercept of participant. To decompose any possible interactions, we conducted follow-up tests using the *interactions* and *emmeans* packages.

A secondary analysis investigated whether children's own racial identity influenced their responses. Due to our relatively small sample sizes, we created 2 broad racial categories. Children were considered "White" if they were monoracial White and "Non-White" if they were a race other than White or multiracial.

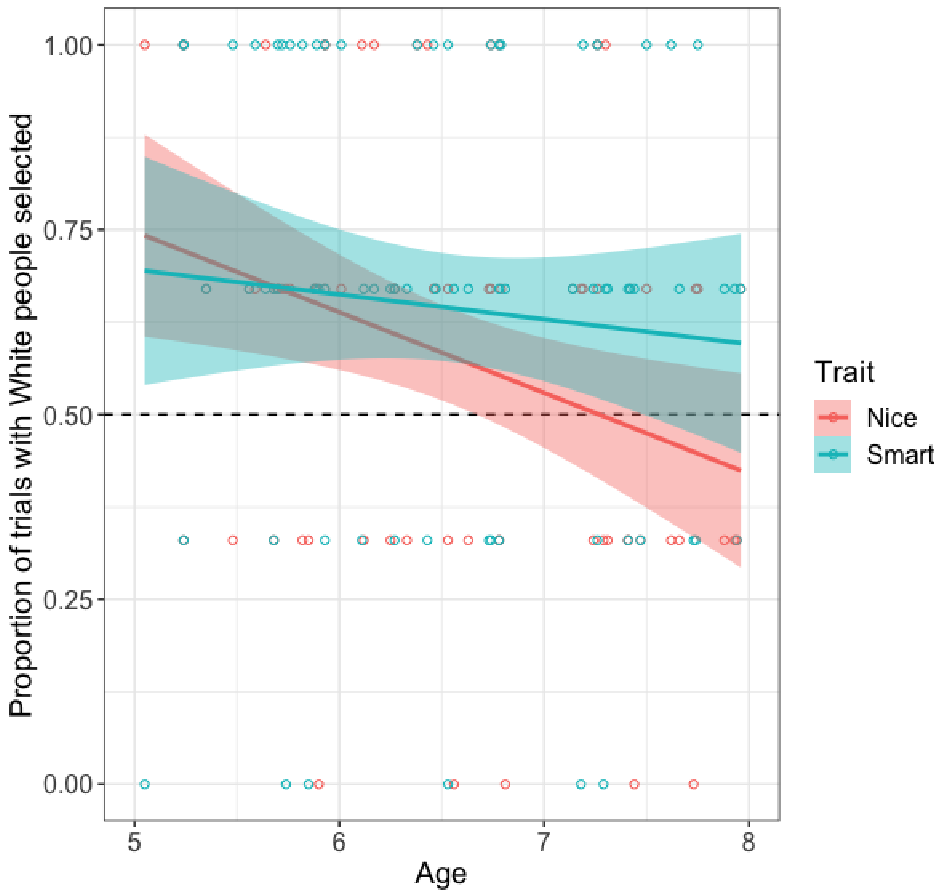
Results and Discussion

We first submitted children's stereotype scores to a linear regression model including trait (nice versus smart), age (continuous), their interaction as predictors, as well as a random intercept for participants. The analyses revealed a main effect of age, B = -0.07, SE = 0.03, t = -2.32, p = .024, suggesting that children became less likely to choose White individuals with age. There was no significant effect of trait, B = 0.07, SE = 0.05, t = 1.35, p = .182. Although the interaction between age and trait did not reach significance, B = 0.08, SE = 0.06, t = 1.25, p = .215, we explored the effect of age in children's stereotypical beliefs about niceness and smartness separately. Simple slope analyses revealed that with age, children became significantly less likely to select White people in nice trials, B = -0.11, SE = 0.04, t = -2.53, p = .013. However, children's tendency to choose White people as being really smart remained stable across age, B = -0.03, SE = 0.04, t = -0.78, p = .437 (Figure 1.1).

Next, we estimated the precise ages at which children's proportions of selecting White people differed significantly from chance (0.5). Following the Johnson-Neyman "regions of significance" approach (Johnson & Neyman, 1936), we subtracted 0.5 (chance level) from children's proportions of selecting White individuals and submitted these scores to the linear regression model described above. Next, we estimated marginal means between ages 5 and 7 by 0.1-year increments. Children's proportion of selecting a White person as being nice was significantly above chance until children reached age 6.6 years. However, children's tendency to select a White person in smart trials remained significantly above chance across age.

Figure 1.1

The association between age and stereotype scores by trait in Experiment 1. The lines show the predicted values from a linear regression model predicting children's stereotype scores from age; the dashed line represents chance; the circles represent the data of individual participants; the shaded areas represent 95% confidence intervals.



Lastly, we tested whether children's own race influenced their responses. We ran a new linear regression model that also included children's own race–White (N = 24) or non-White (N = 27)– as a predictor. The analysis suggests a significant interaction between child race and trait, B = 0.27, SE = 0.11, t = 2.44, p = .019, which was not moderated by child age, B = -0.01, SE = 0.14, t = -0.09, p = 0.928. We then calculated the estimated marginal means across age and conducted pairwise comparisons. Non-White children selected a White person similarly across both trial types, in 69% of "nice" trials and 59% of "smart" trials, t = 1.24, p = .223. In contrast, White children were significantly more likely to select a White person in "smart" trials (71%), than in "nice" trials (53%), $t = -2.18$, $p = .034$ (Figure S1). These findings suggest that children do not simply act on an overall positivity about their own race or a general preference for White people.

The results of Experiment 1 suggest that children in early childhood develop stereotypical beliefs about White and Black peoples' warmth and competence. Specifically, when asked about who is nice, the tendency to select White people over Black people was present among 5- and 6-year-olds. By age 6.7, children became equally likely to choose a White individual or a Black individual as being nice. When asked about who is smart, however, 5- to 7-year-olds were more likely to choose White individuals over Black individuals, suggesting that the stereotype linking intellectual capacities with White people takes root in early childhood. Importantly, these results are not due to a general preference for White people or an in-group racial preference.

Experiment 2: Perceptions of White and Asian Individuals

Experiment 2 examined children's stereotypes about competence and warmth when making judgments of White people and Asian people. Asian people face mixed stereotypes such that they are perceived as more competent but less warm than White people (e.g., Fiske et al., 2002; Lin et al., 2005). If children are sensitive to the content of common racial stereotypes, rather than merely acting on their favoritism towards their ingroup members or White people, they would link Asian people with relatively high competence and low warmth. Thus, Experiment 2 afforded a stronger test of our proposal that children attend to SCM to construct their racial beliefs.

Method

Participants

Our sample consisted of 72 children (24 per age group; 36 girls and 36 boys) aged 5 ($M = 5.49$ years), 6 ($M = 6.63$ years), and 7 ($M = 7.58$ years). Children were recruited from the same database and museum mentioned in Experiment 1. None of them participated in Experiment 1. Children were tested at or near a major university in the Western United States, either in the laboratory ($n = 48$) or a local museum ($n = 24$) and received a small toy as compensation. Data was excluded from an additional 13 children who did not pass the screener task. Caregivers provided racial demographic information for 56 children. Of these children, 55% were White, 23% were multiracial (i.e., selected more than one race), 16% were Asian, 4% were American Indian or Alaska Native, and 2% were Black or African American.

Procedure and Measures

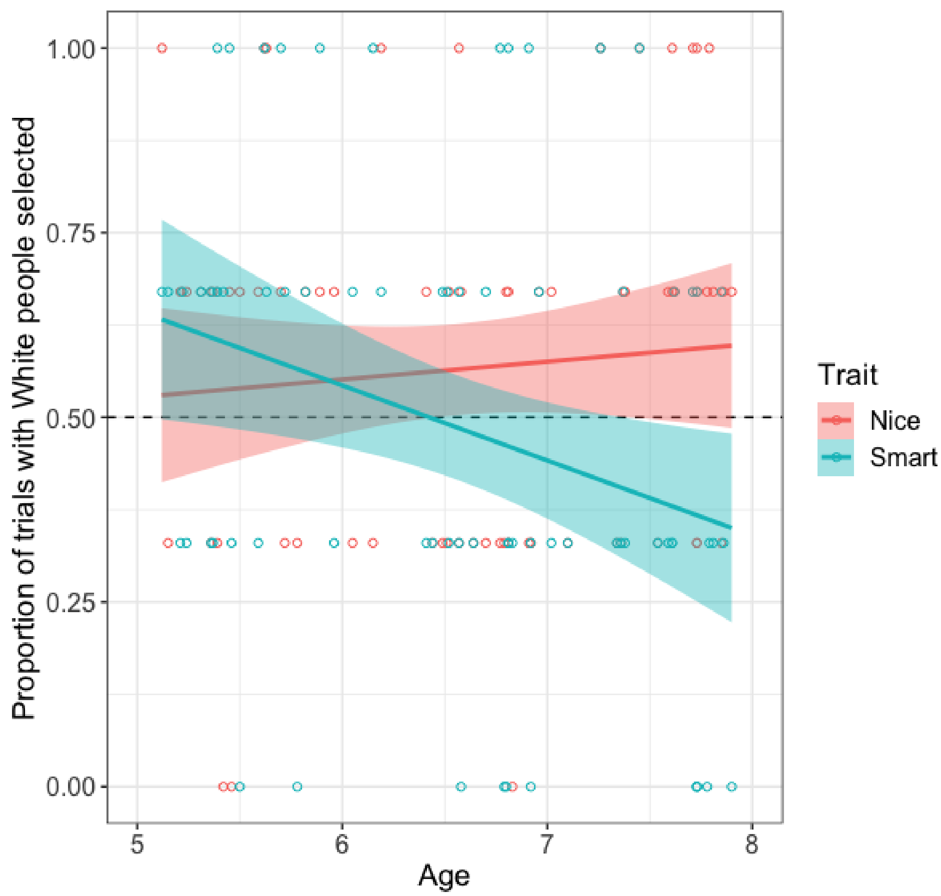
Data collection started in June 2019 and ended in March 2020. The procedure and measures were identical to those used in Experiment 1 except for the stimuli. Here we selected pictures of targets who were judged to be clearly White or Asian, male or female, and we matched average ratings of perceived attractiveness and happiness across race. Each photo used has ratings from an average of 32 raters.

Results and Discussion

As in Experiment 1, we submitted children's stereotype scores to a linear regression model including trait (nice versus smart), age (continuous), their interaction as predictors, as well as a random intercept for participants. There was no main effect of age, $B = -0.04$, $SE = 0.03$, $t = -1.40$, $p = .165$, or trait, $B = -0.08$, $SE = 0.05$, $t = -1.76$, $p = .083$. However, there was a significant age and trait interaction, $B = -0.13$, $SE = 0.05$, $t = -2.48$, $p = .016$. A simple slopes test found that, with age, children became less likely to associate smartness with White people than Asian people, $B = -0.10$, $t = -2.71$, $p = .008$, but their tendency to associate niceness with White people did not vary by age, $B = 0.02$, $t = 0.64$, $p = .521$ (Figure 1.2).

Figure 1.2

The association between age and stereotype scores by trait in Experiment 2. The lines show the predicted values from a linear regression model predicting children's stereotype scores from age; the dashed line represents chance; the circles represent the data of individual participants; the shaded areas represent 95% confidence intervals.



Next, we estimated the precise ages at which children's proportions of selecting White people differed from chance, following the Johnson-Neyman "regions of significance" approach

(Johnson & Neyman, 1936). Children's tendency to select White people as being nice became significantly higher than chance at 6.5 years, whereas their tendency to select White people as being smart became significantly *lower* than chance at age 7.4 years.

We again tested whether children's own race influenced their responses. We ran a new linear regression model that also included children's own race–White (N = 31) or non-White (N = 25)– as a predictor. There was no evidence that child race interacted with the results reported above, *p*s > .250, suggesting that White and non-White children hold similar racial beliefs about competence and warmth when evaluating White and Asian individuals (Figure S2). Given the small sample size, these results should be interpreted with caution.

The results of Experiment 2 suggest that stereotypes about White and Asian people's warmth and competence develop early. Around age 7, children begin to attribute smartness to Asian people and niceness to White people. These findings cannot be explained by children's overall positivity about their own race or White people; rather, children assimilate the fine-grained racial stereotypes about warmth and competence from early on.

Discussion

Across two experiments, we demonstrate that racial stereotypes about warmth and competence emerge in early childhood. In Experiment 1, children as young as 5 associated niceness and smartness with White people more than Black people. The tendency to associate niceness with White people more than Black people decreased with age. However, children's tendency to associate smartness with White people more than Black people remained stable. In Experiment 2, we demonstrate that 5- and 6-year-olds were equally likely to select White and Asian people as being nice and smart, but by the time children reached age 7, they associated niceness with White people more than Asian people, and smartness with Asian people more than White people.

Although previous research has demonstrated that older children and adolescents endorse racial stereotypes (e.g., Cvencek et al., 2015; Pauker et al., 2010), we present the first evidence demonstrating that children as young as 5 to 7 years of age apply the stereotype content model in constructing beliefs about racial categories. Identifying the developmental roots of racial stereotypes is a crucial first step to design and launch interventions at the optimal time point. Past work has shown that children act on stereotypes to guide their behaviors (Bian et al., 2017; Master et al., 2021). For example, children may induce stereotypical behavior through a process of self-fulfilling prophecy, which in turn confirms their stereotypes (Bian et al., 2017; Darley & Fazio, 1980). Children also serve as gatekeepers excluding members of categories that are associated with low competence and thus perpetuate stereotypes (Bian et al., 2018). Thus, devising strategies for young children who have not yet developed strong racial beliefs presents a unique and promising avenue for reducing harmful racial stereotypes.

These findings are in line with previous work that explains the developmental processes that allow children to develop stereotypes. For example, our findings support the idea that children use race as a meaningful dimension by which to categorize people, and subsequently, form stereotypes about different racial groups (DIT; Bigler & Liben, 2007). Also, consistent with SIT

(Tajfel & Turner, 1979) and SRD (Rutland et al., 2010) we found that children's own racial identity slightly moderated their endorsement of racial stereotypes. Extending this past research, we found that children do not simply act on an in-group preference or a preference for White people, but rather, their beliefs are shaped by the SCM. Further, despite previous research with mixed findings, our results provide evidence that children have a fine-grained view of attributes and can differentiate between the dimensions of warmth and competence, two dimensions central to social cognition across culture and time (Fiske, 2018).

How do children acquire the racial stereotypes linking low warmth and low competence with Black people, while linking low warmth and high competence with Asian people? There are at least two possibilities. One is that children are sensitive to socioeconomic status (Olson et al., 2012; Shutts et al., 2013), which may shape their inferences of competence and warmth. In 2019, the poverty rate for Black people was around 19%, and around 7% for Asian and White people (U.S. Census Bureau, 2020). In turn, compared to Asian-American and White students, Black students are more likely to attend high-poverty schools offering fewer math and science courses that prepare students for college (U.S. Government Accountability Office, 2019) and are less likely to earn a college degree (National Center for Education Statistics, 2019). Perhaps children pick up on cues of racial disparities in socioeconomic status and generate stereotypes about competence and warmth (Bigler & Liben, 2007). Indeed, a recent study found that children believed that wealthy individuals were more likely to be smart and less likely to be nice than poor individuals (King & Kinzler, 2022). Another possibility is that children acquire these stereotypes from what they see in the media. In U.S. television, Black characters are typically depicted as less respected and possessing a lower professional status compared to characters of other racial or ethnic groups (Tukachinsky et al., 2015). Similarly, Black people are often portrayed in an aggressive and unfavorable manner in mainstream magazines (Hazell & Clarke, 2008). On the other hand, Asian-Americans are rarely depicted in television, and when they are, these characters tend to fit into the "model minority" stereotype, giving them qualities such as intelligence and social awkwardness (Mastro, 2017). It is also worth noting that children's stereotypes about racial groups do not simply differ in content. Our results demonstrate that the developmental trajectories of their stereotypes vary. Why do children endorse stereotypes about Black people earlier than those about Asian people? We speculate that this is in part due to Asians' invisibility in the U.S. in various aspects including public health, media portrayals and the political world (Mastro, 2017; Muramatsu & Chin, 2022). Further, due to the model minority myth (Museus & Kiang, 2009) Asian people may be omitted from conversations about racial and ethnic diversity, and instead, be portrayed as a homogenous problem-free group that does not face prejudice and discrimination to the extent that other racial minority groups do.

The current research highlights several avenues for future research. First, it is important to consider children's own race when evaluating their endorsement of racial stereotypes. Although racial stereotypes are prevalent and widely endorsed, recent work suggests that children's own racial group membership may affect their reasoning about race (Dautel & Kinzler, 2018; Mandalaywala et al., 2019; Pauker et al., 2010). In Experiment 1, we found slight differences between White and non-White children's racial beliefs about competence and warmth, but these findings should be interpreted with caution because of relatively small sample sizes. Future work should explore this question with high-powered, diverse samples. Second, it will be worthwhile to investigate the development of racial stereotypes as they intersect with other aspects of identity, such as gender or socioeconomic status. Children consider multiple identities when reasoning about social categories (Jaxon et al., 2019; Lei et al., 2020; Perszyk et al., 2019; Shu et al., 2022). For example, 6-year-old children attribute brilliance to White men (versus White women), but they do not apply this gender stereotype to people of racial minority (Jaxon et al., 2019; Shu et al., 2022). Building on this research to explore how children's racial stereotypes intersect with gender is an important next step. Lastly, future work should investigate other aspects of warmth and competence. Our research focused specifically on perceptions of niceness and smartness, two important aspects of warmth and competence, but several other relevant traits exist (e.g., trustworthiness, confidence, etc.). Children's understanding of other more complex characteristics within the two dimensions may have different developmental timelines.

In conclusion, the current research presents evidence suggesting that children are sensitive to the Stereotype Content Model when forming racial stereotypes. Children as young as 5 attributed warmth and competence to White people more than Black people (though the association between warmth and White people decreased with age), and children as young as 7 associated competence with Asian people and warmth with White people. This work sheds light on the early development of harmful racial stereotypes and highlights several potential avenues for future research addressing the current racial disparities.

Chapter 2: Children's Use of Pragmatic Language to Learn About the Social World[1]

Abstract

Young children endorse stereotypes—such as "girls are bad at math." We explore one mechanism through which these beliefs may be transmitted: via pragmatic inference. Specifically, we ask whether preschoolers and adults can learn about an unmentioned social group from what is said about another group, and if this inferential process is sensitive to the context of the utterance. Sixty three- to five-year-old children and fifty-five adults were introduced to two novel social groups —Stripeys and Dotties— and witnessed a speaker praising abilities of one group (e.g., "the Stripeys are good at building chairs"). To examine the effect of context, we compared situations where the speaker was knowledgeable about the abilities of both groups, and had been queried about the performance of both groups (broad context), vs. situations where the speaker was only knowledgeable about one group and was only asked about that group (narrow context). Both preschoolers and adults were sensitive to context: they were more likely to infer that the group not mentioned by the speaker was relatively unskilled, and were more confident about it, in the broad context condition. Our work integrates research in language development and social cognitive development and demonstrates that even young children can "read between the lines," utilizing subtle contextual cues to pick up negative evaluative messages about social groups even from statements that ostensibly do not mention them at all.

Introduction

Young children often endorse stereotypical beliefs about a social group's traits and abilities. For example, by first grade, children associate intellectual ability with men more than women and believe that women are bad at math (Bian et al., 2017; Cvencek et al., 2011; Master et al., 2021). But how are beliefs like these transmitted to children? The present study explores the role of linguistic cues. One possibility is that children learn these stereotypes via testimony, by hearing explicit negative claims such as "the girls are bad at math." However, children may be unlikely to hear such statements from adults, given prior evidence that adults do not typically express

---

[1] This chapter has been published in: Baharloo, R, Vasil, N., Ellwood-Lowe, M.E., & Srinivasan, M. (2023). Children's use of pragmatic inference to learn about the social world. *Developmental Science*, *26*(3), e13333. https://doi.org/10.1111/desc.13333. Supplemental Materials for this chapter are available at the provided link.

biased attitudes explicitly, even if they hold them implicitly (Baron & Banaji, 2006; Dovidio, 2001). Another possibility—explored in the present study—is that children can learn stereotypes from testimony, by reasoning pragmatically about what a speaker could have said about a group but chose not to say (Grice, 1975).

To illustrate, consider the example of a teacher who states, "the boys are good at math." The context of the utterance can dramatically influence the inferences it warrants. If the teacher said this in a classroom full of girls and boys, responding to a general inquiry about children's math ability, a listener would be justified to infer with some confidence that the girls are not good at math ("the boys [but not the girls] are good at math"), based on the teacher's decision to omit the girls and praise the boys only. By contrast, the same statement in response to an inquiry about the boys' performance in a boys-only classroom would not warrant any inferences about the math abilities of girls. Instead, a listener would likely interpret it as highlighting a different contrast, e.g., "the boys are good at math [rather than bad at it, or rather than at geography]"; if asked about the relative math skills of boys vs. girls, the listener should be less confident than in the first scenario. The present study asks whether preschoolers and adults make such context-sensitive pragmatic inferences about unmentioned social groups.

Prior work suggests that children struggle to make some pragmatic inferences until surprisingly late in development. For example, 5- to 9-year-olds fail to recognize that when a speaker uses a weaker term on a scale (e.g., Ernie ate *some* of the cookies), they are unlikely to be expressing a logically compatible, stronger meaning on the scale (e.g., that Ernie ate *all* of the cookies; (Huang & Snedeker, 2009; Noveck, 2001; Papafragou & Musolino, 2003). One possibility is that young children's failure to make these "scalar implicatures" reflects a general sociocognitive difficulty that might also prevent them from making other kinds of pragmatic inferences. However, more recent work suggests that young children may struggle to make scalar implicatures due to a specific difficulty with accessing scalar alternatives from memory (e.g., understanding that a speaker could have said "*all* of the cookies" instead of "*some* of the cookies"), opening the possibility that they could make other kinds of pragmatic inferences, like those of interest here. Indeed, several studies show that when the context makes alternative utterances accessible, even preschoolers compute pragmatic inferences (Barner et al., 2011; Stiller et al., 2015).[2]

Building on this recent work, research suggests not only that preschoolers make pragmatic inferences, but also that pragmatic inference can provide an important tool for learning about the world. For example, preschoolers can identify the typical properties of a category from how a speaker chooses to describe a specific exemplar: e.g., upon hearing "this one is a tall zib," 4.5-year-olds infer that most zibs are shorter than the exemplar under discussion (Horowitz & Frank, 2016). There is also evidence that pragmatic reasoning can support children's learning about

[2] For example, Stiller et al. (2015) presented 3.5 to 4-year-old children with pictures of a person wearing glasses, a person wearing glasses and a hat, and a person not wearing either. After children heard "my friend has glasses," and were asked to guess which person was the speaker's friend, they selected the person wearing the glasses only, inferring that the speaker would have used a more informative expression had they intended to refer to the person with the *glasses and hat*. Children may have succeeded here because the alternative utterances (e.g., *glasses*, *glasses and hat*) were scaffolded by the referential context.

social groups. For example, 7- to 11-year-olds (and adults) infer that one social group is better than another group from claims that explicitly assert similarities, by reasoning about which group the speaker chooses to frame as the reference group: e.g., "girls are as good as boys" implies that boys are generally better (Chestnut et al., 2021; Chestnut & Markman, 2016). Recent research also suggests that hearing generic language from a speaker about one social group ("Zarpies are good at baking pizzas")—as opposed to specific non-generic language about one member of that group ("This Zarpie is good at baking pizza")—leads preschoolers to make inferences about a member of a contrasting social group (e.g., that a Gorp is not good at baking pizzas; Moty & Rhodes, 2021). Although the focus of this prior work was on the effect of generic language in particular, there is reason to think that similar inferences may arise in situations where no generic language is used, as in the example we began with, in which a teacher says that "the boys are good at math" (as opposed to a generic alternative, "boys are good at math"). This issue is important to address because children may be more likely to hear non-generic language about members of a social group than generic language about that group, given prior research suggesting that adults are less likely to produce generic statements about a group when they do not hold essentialist beliefs about that group (Rhodes et al., 2012).

Here we build on this prior literature to investigate preschoolers' sensitivity to contextual cues in their learning from testimony about social groups. Specifically, we examine whether 3- to 5-year-olds and adults can learn about a novel social group (introduced as an alien group on another planet) from what a speaker says about another social group, and whether this inferential process is sensitive to the context of the utterance. Participants were assigned to either a *broad* or *narrow* condition. In both conditions, participants heard a speaker attribute a property to members of one group—e.g., that "the Stripeys are pretty good at building chairs". What varied critically between the conditions was the speaker's knowledge about the abilities of members of the second group, and the question that the speaker was answering. In the *broad condition*, the speaker was knowledgeable about both groups and the question concerned both groups (e.g., "So, what do you think of the chairs the Stripeys made and the chairs the Dotties made?"). But in the *narrow condition*, the speaker was only knowledgeable about one group and was only asked about that group (e.g., "So, what do you think of the chairs the Stripeys made?").

Of primary interest was what children (and adults, for comparison) would infer about the group that was not explicitly mentioned by the speaker; specifically, whether that group would be perceived as relatively more or less skilled than the mentioned group. We expected that in both contexts children and adults would learn a new fact about the high skill level of the mentioned group (e.g., that the Stripeys are pretty good at building chairs). But, if participants are sensitive to the context, their inferences about the properties of the unmentioned group (e.g., the Dotties) should vary by condition. In the broad context, the speaker's omission of one group likely conveys a lower opinion about the unmentioned group's skill level. We thus expected that in the broad context, children and adults would indicate that the unmentioned group was less skilled at the target ability than the mentioned group, and feel relatively confident in their judgment. In the narrow condition, however, the context itself might account for the narrow referential scope of the utterance: the speaker may have only praised one group because they were only asked about that group. Thus, in the narrow condition, we expected to see a less pronounced differentiation in participants' evaluations of the skill levels of the mentioned vs. unmentioned groups, and lower confidence in their assessment.

Method

Participants

We tested 60 3-5-year-old children (*M* = 4.44 years, *SD* = 0.7 years, 26 females, 34 males) and 55 adults (*M* = 22.3 years, *SD* = 3.9 years, 34 females, 21 males). Participants in each age group were randomly assigned to the broad condition (30 children, *M* = 4.53 years; 28 adults) or the narrow condition (30 children, *M* = 4.36 years; 27 adults). The target sample size of 30 children per condition was based on the sample sizes of studies of pragmatic inference in children (e.g., Stiller et al., 2015). Participants were recruited and tested at or near a major university in a racially diverse but majority white area in the Western United States. Children were recruited from a database of families who were interested in research opportunities, as well as local museums and preschools, and were tested in the laboratory or at the recruitment site and compensated with a small toy. Data from 11 children were excluded due to experimenter errors (*n* = 4), video-recording issues (*n* = 3), wrong or unknown age (*n* = 2), or sibling/caregiver interference (*n* = 2). Data from 1 adult were excluded due to video-recording issues. Adults were tested in the laboratory and participated in exchange for course credit or snacks. Data were collected from July 2017 to June 2019 (children) and June 2018 to March 2019 (adults). This study was approved by the Institutional Review Board at the University of California, Berkeley. Parents gave informed consent for their children and adult participants gave informed consent for themselves.

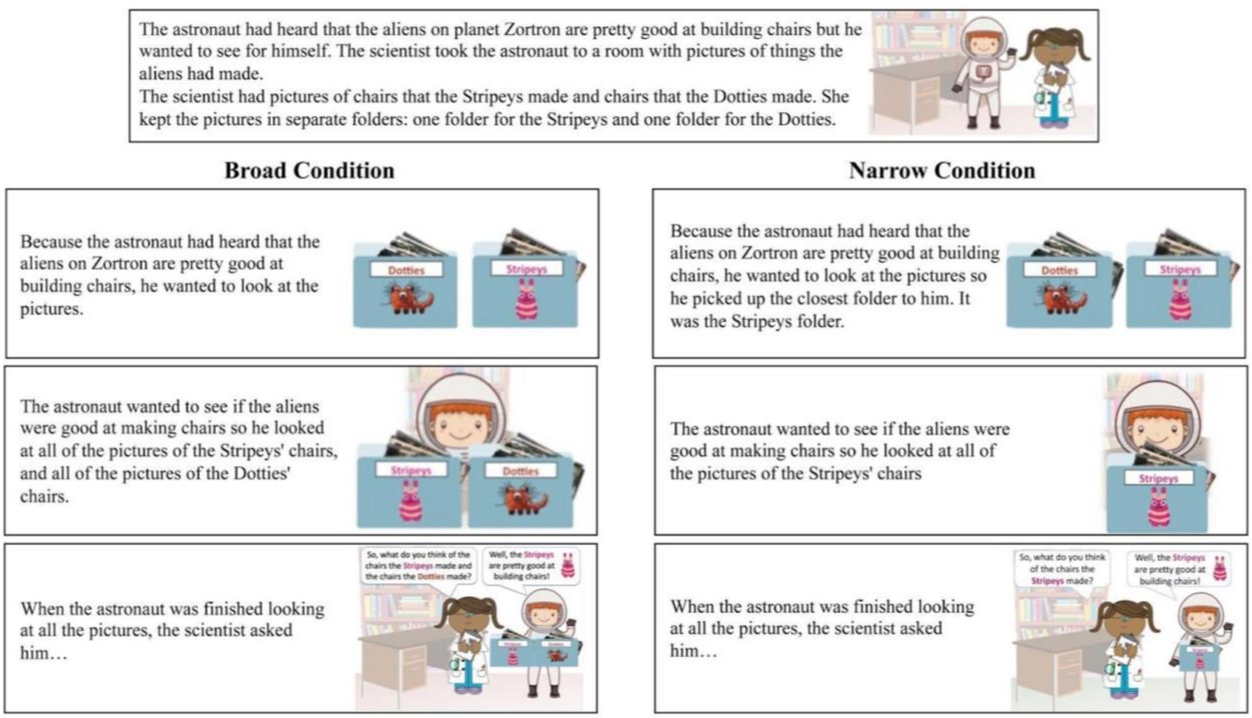
Design and Procedure

An experimenter led each participant through an illustrated story involving two novel alien groups—Stripeys and Dotties—who differed in physical appearance (see Figure 2.1). Two vignettes described the aliens' ability at building chairs or sandcastles (order counterbalanced). At the end of each vignette, participants heard a statement about *one* of the alien group's abilities, referencing different groups across the two trials (e.g., "Well, the Stripeys are pretty good at building chairs!" and "Well, the Dotties are pretty good at building sandcastles!"). Participants then indicated how good they thought the objects made by each group were, which alien group had better building abilities (or if they were the same), and how sure they were in their judgment of which group was better. The full experimental protocol, raw data files, and analysis scripts can be found on OSF at https://osf.io/ceq84.

Materials

The story introduced a scientist studying Stripeys and Dotties on planet Zortron. The scientists had taken photographs of some chairs and sandcastles that the aliens built and showed them to a visiting astronaut who had heard that the aliens are pretty good at these skills, but "wanted to see for himself". In the **narrow context** condition, the astronaut happened to see objects made by one group only: he randomly picked up the photo folder "closest to him", and it happened to contain either the Stripey- or Dottie-made chairs / sandcastles. When the astronaut was finished looking through that group's photo folder, the scientist probed the astronaut's opinion specifically about that group's skill, e.g., "So, what do you think of the chairs that the Stripeys made?". In contrast, in the **broad context** condition, the astronaut ended up looking at objects made by both groups, e.g. "he looked at all of the pictures of the Stripeys' chairs, all of the pictures of the Dotties' chairs". The scientist then asked the astronaut's opinion about the skill of both alien groups: “So, what do you think of the chairs the Stripeys made and the chairs the Dotties made?”. In sum, the critical manipulation of narrow vs. broad context concerned whether the astronaut had knowledge of one or both groups’ abilities (based on whether he had seen photos of the objects they built), and whether the question under discussion concerned one or both groups (Figure 2.1). Finally, in both narrow and broad contexts, the astronaut responded with the exact same utterance: “Well, the Stripeys are pretty good at building chairs!”[3]

Figure 2.1
Sample trial.



Measures

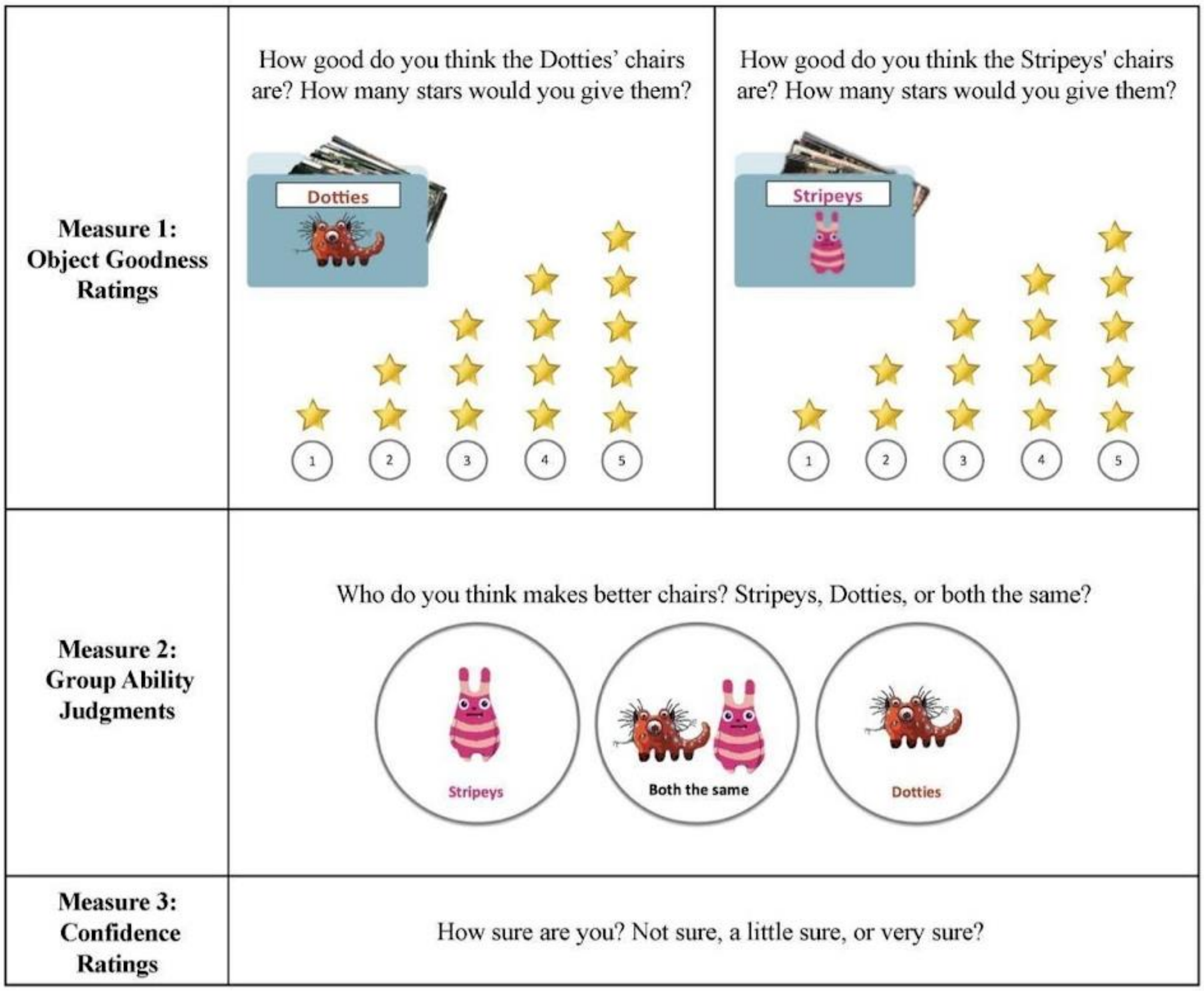
Participants were first familiarized with the story characters and objects relevant to the study and were introduced to the rating scale used in the test trials (Supplementary Material). Participants

[3] Generally, in the absence of any evidence about a group, a rational guess is that they will have “average” skills. We reasoned that this assumption might drive apart skill ratings for the mentioned and unmentioned groups in *both* the narrow and broad conditions, since hearing that members of the mentioned group are “pretty good” would suggest that their skill is above average to a noteworthy extent. To mitigate this possibility, participants in both conditions learned that--prior to examining the photo evidence--the astronaut already had a baseline belief that the aliens’ skills were fairly high (“The astronaut had heard the aliens on planet Zortron are pretty good at building [chairs / sand castles]”). Setting up this elevated baseline expectation allowed us to pinpoint the potential downgrading of beliefs about the unmentioned group specifically due to the pragmatic context of the utterance. As an additional way to ensure a level playing field for the unmentioned group in every way other than the pragmatic context, in the broad condition we specified that the astronaut chose to look at the folder he did out of convenience, because it was the closest one to him (as opposed to picking the folder because he thought that group’s abilities were high in particular).

were also asked to identify the story characters from pictures (astronaut, scientist, Stripeys, and Dotties), and received feedback if needed.

After each trial, participants were asked to provide a series of judgments in the following order (Figure 2.2).

Figure 2.2
Sample visual aid for the three judgments in the chair trial.



Measure 1: Object Goodness Ratings
After each trial, participants rated how good the objects (chairs or sandcastles) made by each group were, using a 5-point star rating scale: "How good do you think the [Dotties' / Stripeys'] [chairs / sandcastles] are? How many stars would you give them?" (Figure 2.2). This measure allowed participants to provide a fine-grained judgment about each of the groups' abilities, one-by-one, on an absolute level. All participants were given practice using this scale at the beginning of the study (Supplementary Material).

Measure 2: Group Ability Judgments
Next, participants were asked to indicate which alien group was better at the target ability, in general. Participants were presented with pictures of the aliens and were asked "Who do you think makes better chairs [sandcastles]? Stripeys, Dotties, or Both the same?" (Figure 2.2). This forced-choice measure makes the comparison between groups salient and asks participants for a relative judgment. It also allowed us to follow-up with a question about participants' confidence in their relative judgment (Measure 3).

Measure 3: Confidence Ratings
Finally, participants were asked to indicate how confident they were in their judgment of which group was better at the target ability. Specifically, they were asked "How sure are you? Not sure, a little sure, or very sure?

Results

Measure 1: Object Goodness Ratings
Table 2.1 shows mean object goodness star ratings and Figure 2.3 shows the distribution of ratings. Participants' ratings were treated as an ordinal variable with five levels (with '5' indicating the highest level) and were predicted from context (narrow vs. broad), target group (mentioned vs. unmentioned), participant age group (child vs. adult), and their interactions, allowing for random participant intercepts, in an ordinal regression with flexible thresholds, using the *clmm* command in R. This analysis revealed a significant three-way interaction, likelihood ratio test $\chi2$ (1) = 5.47, p = .019. We explored this interaction further by analyzing the child and adult data separately. The interaction between target group and context was significant for children, $\chi2$ (1) = 10.27, p = .001, but not adults, $\chi2$ (1) = 3.67, p = .055.

Table 2.1
Means and Standard Deviations for Children's and Adults' Object Goodness Ratings in the Broad and Narrow Contexts

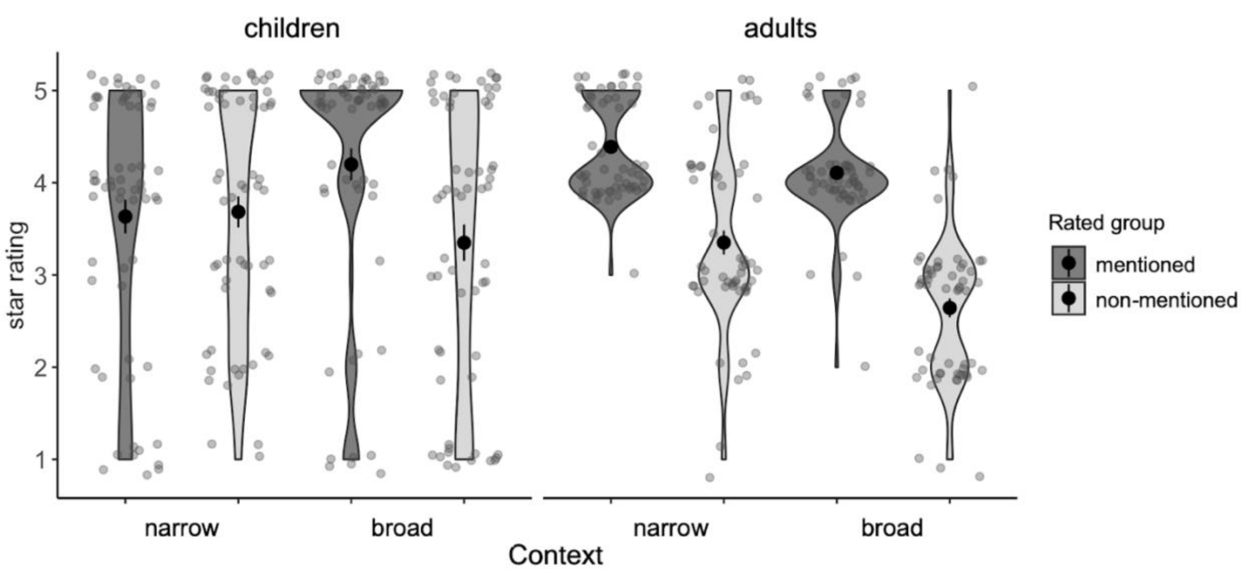
Age group	Broad Context		Narrow Context	
	Mentioned Group	Unmentioned Group	Mentioned Group	Unmentioned Group
Children	4.2 (1.35)	3.35 (1.54)	3.63 (1.41)	3.68 (1.31)
Adults	4.11 (0.56)	2.64 (0.77)	4.39 (0.52)	3.35 (0.96)

***Note.* Values in parentheses are standard deviations.**

As shown in Figure 2.3, the predicted pattern emerged in both age groups: in the broad context condition, where the astronaut ostensibly chose to selectively praise one group only—even though he was queried about both groups and possessed the relevant first-hand knowledge to evaluate them—participants in both age groups rated the objects made by the unmentioned group as less good than those made by the mentioned group (children: $z = -3.96$, $p < .001$; adults: $z = -8.47$, $p < .001$). In the narrow context condition—where the astronaut was queried about one group only, and possessed relevant first-hand knowledge only about that group—children did not take the astronaut's positive comment to imply that the unmentioned group's objects were any worse than the mentioned group's objects, $z = .26$, $p = .793$. Adults, however, did rate the unmentioned alien group's objects lower than the mentioned alien group's objects even in the narrow condition, $z = -6.85$, $p < .001$, presumably because they had received evidence about the high level of ability of the mentioned group but no evidence regarding the unmentioned group; still, this effect was weaker than in the broad condition.

Figure 2.3

Distribution of object goodness ratings as a function of age group (children vs. adults), context (narrow vs. broad) and rated group (mentioned vs. non-mentioned by the speaker). Raw ratings are plotted as jittered datapoints. Cell means and standard errors (shown as black dots with error bars, ±1SE) are displayed for reference only; the analyses treated star ratings as an ordinal variable.



Additional comparisons revealed that children rated the mentioned alien group's objects higher in the broad than narrow condition, $z = 2.78$, $p = .005$, but their ratings of the unmentioned alien group's objects did not differ significantly across conditions, $z = -1.16$, $p = .244$. Adults' object ratings were lower in the broad than narrow condition, both for the mentioned alien group, $z = -1.97$, $p = .049$, and the unmentioned alien group, $z = -3.85$, $p < .001$.

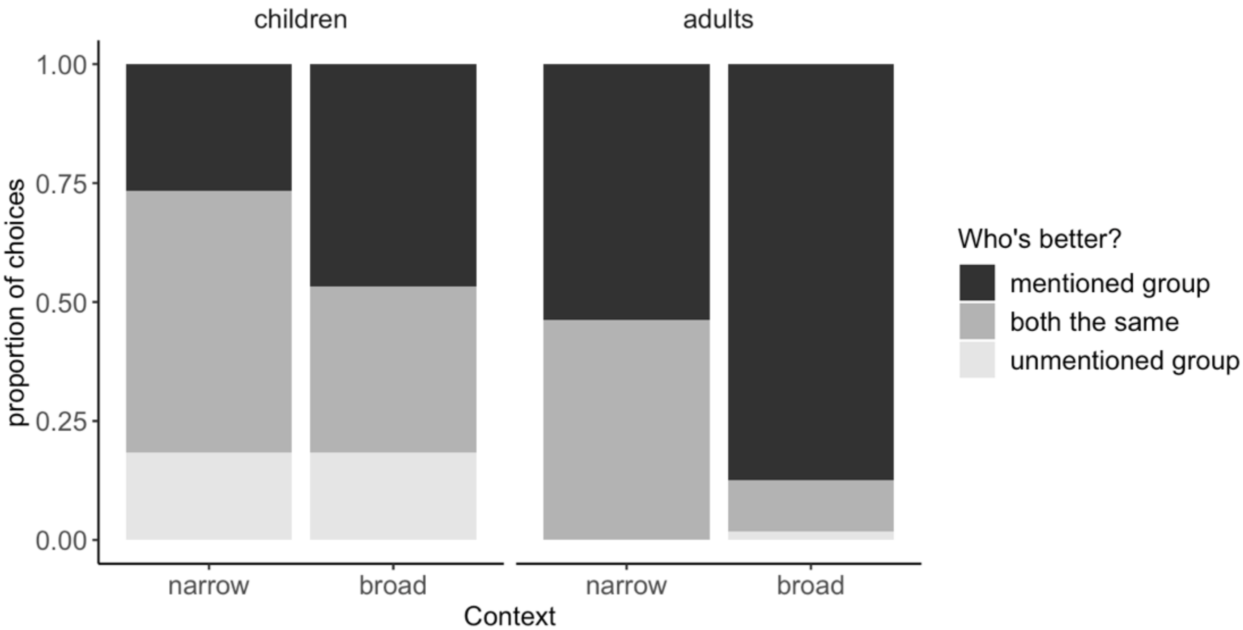
As Table 2.1 indicates, children in the broad condition rated both the mentioned group (*M* = 4.2) and unmentioned group (*M* = 3.35) above the scale midpoint (3); ratings for each group hovered just above the midpoint in the narrow condition (mentioned: 3.63; unmentioned: 3.68). Thus, children rated the groups above average across conditions, such that the effect was driven by the higher rating given to the mentioned group in the broad condition, as opposed to an overall negative evaluation of the unmentioned group. This could be taken to imply that the children did not infer, in the broad condition, that the unmentioned group was "bad" at producing the objects in question, but only not as good as the mentioned group: i.e., they may not have developed a negative belief about the group's abilities, but only a less positive belief. Still, it is worth noting that we introduced the scale to participants in a way such that they may have believed that even an average rating on the scale corresponded to a poor-quality object: i.e., participants were told that '1' star should be given to something that is poorly done, like a cupcake that was made out of mud and feathers (see Supplementary Material). In contrast to in children, the group x context interaction in adults (which was marginal at $p = .055$) was driven by adults' relatively more negative evaluation of the unmentioned group in the broad condition, whose objects were rated below the scale midpoint (*M* = 2.64; see Table 2.1).

Measure 2: Group Ability Judgments

Figure 2.4 depicts participant's responses to the question of which alien group was better at the target ability. Participants' selections were coded into a 3-point ordinal scale with the highest value reflecting the response that the mentioned group is more skilled (1 = the unmentioned group is better, 2 = both the same, 3 = the mentioned group is better). An ordinal regression predicting choices from context (broad vs. narrow), participant age group (children vs. adults) and their interaction, allowing for random participant intercepts and flexible thresholds, was implemented using the *clmm* command in R. The interaction was not significant, $\chi 2\ (1) = 2.50$, $p = .114$, and was dropped from the model.

Both the context manipulation, $z = -3.15$, $p = .002$, and age group, $z = -4.79$, $p < .001$, were significant predictors of participants' choices. As shown in Figure 4, the context manipulation produced the predicted effect: participants in the broad condition of both age groups were more likely to judge that the mentioned group was better at the target skill than in the narrow condition. Additionally, adults were overall more likely than children to select the mentioned group in this task. Indeed, even in the narrow condition, adults indicated that the mentioned group was better about half of the time (Figure 2.4).

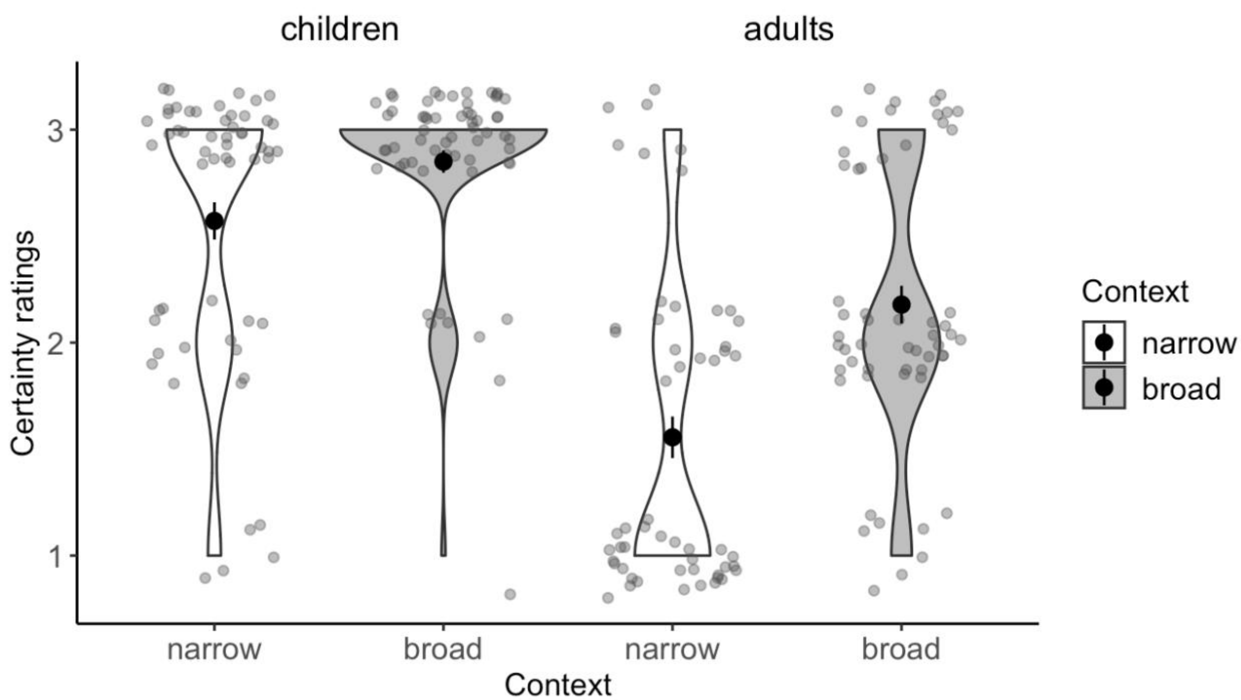
Figure 2.4
Distribution of participants' choices in response to which group was better at the skill in question, as a function of age group (children vs. adults) and context (narrow vs. broad).



Measure 3: Confidence Ratings
Figure 2.5 depicts participants' ratings of how sure they were regarding which group was better at the skill in question. Participant's ratings were treated as a 3-point ordinal scale tracking increasing levels of confidence (1 = not sure, 2 = a little sure, 3 = very sure). An ordinal regression predicting confidence ratings from context (broad vs. narrow), participant age group (children vs. adults), and their interaction, allowing for random participant intercepts and flexible thresholds, was implemented using the *clmm* command in R. The interaction was not significant, $\chi 2$ (1) = .47 , p = .493, and was dropped from the model. Both the context manipulation, z = -3.73, $p < .001$, and age group, z = 5.41, $p < .001$, significantly predicted confidence ratings. Figure 2.5 shows that, as predicted, participants in both age groups were more confident in the broad condition than in the narrow condition. Overall, children also gave higher confidence ratings than adults, consistent with prior research suggesting that young children often exhibit overconfidence in their own knowledge and beliefs (e.g., Hagá & Olson, 2017).

Figure 2.5

Distribution of confidence ratings as a function of context (narrow vs. broad) and age group (children vs. adults). Raw ratings are plotted as jittered datapoints. Cell means and standard errors (shown as black dots with error bars, ±1SE) are displayed for reference only; the analyses treated confidence ratings as an ordinal variable.



Discussion

The present study asked whether three- to five-year-olds and adults can learn about an unmentioned social group from what has been said about another group, and whether this inferential process is sensitive to the context of the utterance. In a situation where a speaker could easily have praised both groups' abilities—because he had knowledge of and was asked about both groups—children and adults inferred that his choice to praise only one group implied that the unmentioned group was less skilled than the mentioned group. This inference was apparent across a range of measures, including participants' ratings of the quality of the objects made by each group, their judgment of which group was better at the ability, and their confidence regarding which group was better. In contrast, when the speaker produced the same utterance in a context where he only had firsthand knowledge of and was only asked about the mentioned group, participants did not draw the same relative inferences about the different skill levels of the two groups. Together, our findings suggest that pragmatic inference may be a mechanism through which children learn stereotypes about social groups, extending previous research on children's developing pragmatic ability (e.g., Barner et al., 2011; Stiller et al., 2015), as well as their ability to learn about the world from language (Chestnut & Markman, 2016, 2018; Horowitz & Frank, 2016; Moty & Rhodes, 2021).

Although the predicted pattern of results emerged for both children and adults across all our three measures, we also found some developmental differences. In line with our hypotheses, children and adults rated the objects produced by the unmentioned group as less good than those produced by the mentioned group, and indicated that the mentioned group was more skilled than the unmentioned group, and these effects were larger in the broad condition than the narrow condition. Notably, however, adults—but not children—also drew inferences about the relative skill level of the two alien groups in the narrow context condition, both in their object goodness ratings and judgments of which group was better. Specifically, adults still rated the unmentioned group as producing lower-quality objects in the narrow condition.

Why did adults, but not children, rate the mentioned group higher than the unmentioned group even in the narrow condition? We had expected that participants might make this inference. For example, if the astronaut praised the Stripeys' chairs, participants may infer that there must be something noteworthy about those chairs (i.e., they are above average). On the other hand, in absence of evidence about the Dotties' chairs, they might just assume that their chairs are of average quality. More generally, in both conditions, participants only received evidence about the high abilities of one group, and were then asked to make judgments both about this group and another group, so it makes sense that they might not have given as high a rating for a group that they received no evidence for. Although we took measures to minimize this possibility (e.g., by trying to establish a high prior baseline for both groups' abilities in the lead-up story; Footnote 3), adults (but not children) still rated the unmentioned group as producing lower-quality objects in the narrow condition.

This opens two possibilities. First, it is possible that both children and adults can make the kinds of inferences that would have led to higher ratings for the mentioned vs unmentioned group in the narrow condition but the measures we took to block these inferences were effective for children but not adults. Or, perhaps adults are more likely to make these kinds of inferences than children, and the measures we took to block them were not strong enough. We favor the second explanation, given adults' more sophisticated pragmatic reasoning. Regardless, we note that even though adults made inferences about the unmentioned group in the narrow condition, we observed stronger effects among adults in the broad condition for all three measures.

Our results converge with and add to prior work showing that hearing generic language from a speaker about one social group leads preschoolers to make inferences about a member of a contrasting social group (Moty & Rhodes, 2021). Most notably, our results suggest that children make such inferences not only in response to generic language, but also in response to non-generic, specific language. This is significant because children may often be more likely to hear speakers use non-generic language in reference to social group members, particularly when speakers do not essentialize the groups in question (Rhodes et al., 2012). Another point of contact between our study and that of Moty and Rhodes (2021) is in establishing the boundary conditions on when children will make an inference about an unmentioned group from a statement about another group. Moty & Rhodes (2021) showed that children generally did not make such inferences when the statement was made by a speaker who was introduced as being from out of town. Yet it is unclear from this manipulation whether children didn't make the inference because they assumed that the out-of-town speaker lacked knowledge of the social groups in question (as the authors suggest), or simply because children assumed that the speaker was unconventional and did not adhere to conversational principles (Grodner & Sedivy, 2011). Our study aimed to more directly manipulate the speaker's knowledge of the groups' abilities by varying whether he had observed pictures of the objects made by both groups (broad condition) or of one group (narrow condition). This difference–combined with our manipulation of whether the question under discussion concerned both groups or just one group--was sufficient to produce different inferences across the two conditions.

Taken together, our findings demonstrate that even preschoolers make contextually sensitive inferences from utterances about social groups, lending plausibility to the idea that these inferences may shape children's beliefs about the social world. One possibility is that the inferences documented here do not only apply to contexts involving contrasting social groups (e.g., such that saying that Stripeys are good at building chairs implies that Dotties are not), but also to contexts with contrasting properties (e.g., such that saying that Stripeys are good at building chairs implies that they are not good at baking). Additionally, it is possible that a similar inferential process may apply outside of contexts in which social groups have been explicitly contrasted. People learn to habitually contrast some groups (like boys and girls), and properties (like science and art), raising the possibility that even when there isn't a strong supporting context, a statement about one group or property (e.g., that boys are strong, or that girls are good at art) could imply that the same is not true of a "chronically-salient," contrasting social group or property (e.g., girls are not strong; girls are not good at science).

In all of the cases described above, children may make a social inference about an unmentioned group or property, even when the speaker did not intend to convey this message. Yet in some situations, leaving something out from an utterance may be the most effective or appropriate way to communicate an intended message. Speakers are expected to adhere to certain norms when speaking, including to be polite and accurate, and societal anti-negativity norms dissuade speakers from openly criticizing individuals or groups. A speaker can adhere to these norms—but still communicate negative sentiments about a group that they wish to convey—by triggering the intended inference with an accurate and positive assessment of a different group (as studied here), or of the same group in a different domain (e.g., describing a group as "passionate" as opposed to "intelligent" in order to convey that the group is ignorant). Bergsieker and colleagues (2012) showed that when participants were asked to describe a person they had just learned about (e.g., who was both intelligent and unkind), they selectively emphasized positive content and omitted negative content (e.g., describing the person as "smart" as opposed to "smart but mean"), and this tendency was more pronounced in a public context. Given popular maxims such as "If you don't have anything nice to say, don't say anything at all," triggering a contrastive inference with a positive assessment may be a socially acceptable way to communicate relatively negative information about social groups. Our study identifies one key ingredient of such "stereotyping by omission": contextual cues that allow the listener to understand that the speaker could have praised a group but chose not to.

In sum, our findings add to a growing body of literature that demonstrates children's ability to use pragmatic inference to learn about the world. We add an important piece to the picture by showing that preschoolers make inferences about unmentioned social groups and documenting the crucial role that context plays in this process. If a speaker could have praised a group, but instead chooses to only praise another group, children use this omission to infer a relatively negative evaluation of the unmentioned group compared to the mentioned group. Arguably, this could be an important mechanism for how children learn negative stereotypes about social groups, which are unlikely to be directly expressed by adults. Our study thus represents an important link between two bodies of literature: the study of pragmatic development, and research on social learning.

Chapter 3: Children Infer the Relative Competence of Individuals by Observing How They Are Addressed

Abstract

Language is a powerful mechanism for transmitting stereotypes. This research explores whether observers' perceptions of a social group's competence are shaped by how that group is addressed. Children aged 5 to 9 and adults observed interactions between a character and two novel social groups, where one group consistently received more complex language, including rarer, later-acquired words with adult-directed prosody (Study 1), or more complex information (Study 2). Results revealed that while adults and older children (8-9 years) attributed higher competence to individuals addressed with sophisticated language and information, younger children (5-6 years) did not. Intriguingly, Study 3 suggested that the younger children's lack of competence-based inferences is not due to an inability to discern content complexity and may instead reflect their developing capacity to associate linguistic complexity with competence. Collectively, these studies underscore how observations of third-party interactions can indirectly shape the development of stereotypes, revealing a subtle yet significant aspect of stereotype transmission.

Introduction

Early Emergence of Stereotypes and Their Downstream Consequences

Understanding the development of stereotypes, particularly those related to competence, is crucial for addressing social inequity. These stereotypes, formed in early childhood, shape individual perceptions and choices, reinforcing systemic biases within society. For instance, by 6 years of age, children often associate brilliance with males more than females, leading girls to avoid activities perceived as requiring high levels of intelligence (Bian et al., 2017). Gender is just one example; stereotypes significantly influence perceptions of groups based on ethnicity, occupation, socioeconomic status, ability, and age (Fiske et al., 2002). These beliefs, firmly established by adulthood, influence personal and professional decisions, perpetuating societal inequalities across diverse groups (e.g., Leslie et al., 2015).

Acquisition of Stereotypes in Children

The question then arises: how do children acquire these stereotypes? This process is influenced by a complex array of factors, with growing interest in the role of subtle, indirect cues. Even as mere observers, individuals' perceptions and biases are shaped by the subtle dynamics they witness in social exchanges (Castelli et al., 2008; Weisbuch et al., 2009). In societies like the United States, where overt bias and prejudice are socially and legally discouraged, the role of these subtle, implicit cues becomes increasingly significant (Baron & Banaji, 2006; Dovidio, 2001). This paper focuses on how subtle cues, particularly in the form of verbal communication, shape observers' perceptions of an individual's competence. For example, a teacher consistently using simpler vocabulary with female students might indirectly and unintentionally convey negative assumptions about the competence of females. We investigate how such variations in communication—observed by children and adults—can subtly yet profoundly influence the formation of stereotypes.

Linguistic Elements and Stereotype Formation

Various linguistic elements, such as group labels, generic statements, omissions, and syntactic framing, subtly but significantly impact the development of stereotypes. Seemingly neutral statements using group labels like "Good morning, boys and girls," may unintentionally emphasize gender divisions, making these categories more psychologically salient to children (Bigler & Liben, 2007). Additionally, generic expressions such as "Boys are good at math" can lead children from a young age to generalize traits across entire groups (Cimpian, 2013; Cimpian & Markman, 2008, 2011; S. A. Gelman et al., 2008; S. A. Gelman & Raman, 2003). While these generalizations might seem harmless, they can have unintended negative consequences. For instance, children might deduce from "boys are good at math" that girls are not, due to the omission of girls from the statement (Moty & Rhodes, 2021). Baharloo et al. (2023) further illustrate that children can form negative inferences about unmentioned groups even when specific statements are made about one group. Moreover, syntactic framing can perpetuate stereotypes by implying a normative comparison, leading even well-intentioned egalitarian phrases like "Girls are as good as boys at math," to inadvertently reinforce stereotypes (Chestnut et al., 2021; Chestnut & Markman, 2016, 2018).

Focus on First-Person Communication

It is well-established that third-person statements–encompassing both what is explicitly said about social groups and what is notably omitted–significantly influence beliefs about these groups. Our research shifts focus to the less explored domain of first-person communication: the direct speech addressed to individuals within social groups. Our studies delve into the nuances of direct communication, analyzing elements such as language complexity, which includes factors like word choice and prosody, as well as content complexity, characterized by its novelty and intricacy. By examining first-person communication, we aim to understand how observers, particularly children, infer beliefs about an individual's competence based on how they are addressed.

Language Adjustments Based on Perceived Audience Competence

Communication often involves speakers adjusting their language based on their audience's perceived knowledge and competence. Clark (1998) highlighted this through the concept of "common ground," wherein speakers modify their language to align with what they believe is mutually understood with the listener. However, this adaptive process can be subtly influenced by stereotypes, particularly those relating to competence. For example, speakers might unconsciously oversimplify their language due to stereotypical beliefs about a listener's knowledge, leading to unnecessary simplifications. Similarly, the process of accommodation—deliberately adjusting speech to align or distance oneself from the listener—can unintentionally reflect stereotypes. For instance, a younger individual may oversimplify their language when speaking to an older person, inadvertently reflecting stereotypes that associate elderly people with lower competence (Giles & Ogay, 2007). Dupree & Fiske (2019) further highlight the pervasive influence of stereotypes in language use. They revealed that White liberals, in aiming to affiliate, may unconsciously rely on negative competence stereotypes about Black individuals, leading them use fewer competence-related words when conversing with Black people compared to White people. These examples illustrate how perceptions of an audience's competence can subtly but significantly influence language use. Building on these insights, our research explores how children observe and interpret linguistic adjustments in first-person communication, focusing on the implications of these observations in the formation of stereotypes.

Children's Sensitivity to Social and Linguistic Cues

Children possess a remarkable ability to interpret subtle or indirect cues in social interactions, a skill that is fundamental to their ability to infer stereotypes and develop biases. Research by (Castelli et al., 2008) demonstrated this proficiency; children formed biases by detecting non-verbal discomfort in interracial interactions, without any explicit communication of prejudice. This sensitivity extends to interpreting emotional expressions to guide their inferences about an individual's competence. Asaba et al., (2020) found that children aged 6 to 9 could assess a student's competence based on a teacher's surprised reaction to their success. If the teacher showed surprise at a student's achievement, children inferred that the student was generally not expected to perform well, thus perceiving them as less competent. These findings illustrate children's sensitivity to social cues, conceptually motivating our research into how children perceive and process the subtleties of first-person communication. This exploration is particularly pertinent in understanding the formation of competence stereotypes among children.

Children's sensitivity to social interactions extends to language, reflecting a deep understanding that language is tailored according to the listener. Toddlers as young as 20 months understand that a speaker's linguistic register is dependent on the listener. For instance, they are surprised when an adult is spoken to in infant-directed speech, but not adult-directed speech (Ikeda et al., 2018). This awareness becomes more pronounced as children grow. By the age of 3, they not only understand but also replicate this audience-specific language adaptation, using simpler phrases when speaking to younger children (Sachs & Devin, 1976; Shatz & Gelman, 1973). This sensitivity extends beyond immediate social interactions: By the age of 4, children begin to infer the identity of an unseen interlocutor based on a speaker's register, demonstrating their ability to understand broader social dynamics through linguistic cues (Wagner et al., 2010).

Current Studies

Our research adopts a novel approach in exploring stereotype development by examining how child and adult observers' perceptions of a social group's competence are influenced by the manner in which the group is addressed. Study 1 delves into the impact of linguistic complexity, including aspects such as lexical choice and prosody. Study 2 shifts focus to the content of communication, investigating how the nature of information—ranging from mundane to intriguing—affects assumptions about a listener's competence. Both studies also explore key aspects of stereotype formation, like the development of group-level perceptions and the extension of these perceptions to new group members. Study 3 then examines the underlying mechanisms driving these inferences, and sheds light on why younger children struggle to make these inferences. We hypothesized that the ability to draw inferences about competence from language cues requires specific linguistic, social, and cognitive skills, including the recognition of variations in how speakers address different individuals and linking these variations to social judgments about competence. As such, we expected that older children (ages 8-9) and adults will be able to perform this complex inferential process more readily than younger children (ages 5 and 6).

All three studies were approved by the Institutional Review Board at the University of California, Berkeley. For each study, parents provided informed consent for their children's participation, and adult participants gave informed consent for themselves.

Study 1: Impact of Linguistic Complexity on Perceptions of Competence

Study 1 investigated whether variations in a speaker's linguistic complexity, specifically lexical choice and prosody, affect children's and adults' perceptions of the listener's competence.

Method

Participants

We tested 96 participants: 32 5-6-year-olds (M = 5.72 years, SD = 0.58 years, 16 females, 16 males), 32 8-9-year-olds (M = 9.01 years, SD = 0.67 years, 16 females, 16 males), and 32 adults (M = 26.68 years, SD = 9.93 years, 16 females, 16 males). Participants were recruited from communities surrounding a large university in the Western United States. Children were recruited from a database of families who were interested in research opportunities, as well as a local museum and preschool, and were tested in the laboratory, online via Zoom, or at the recruitment site. Children were compensated with a small toy or a certificate. Adults were tested in the laboratory, a local park, on the University campus, or virtually via Zoom, and received either course credit or a snack.

We obtained race information for 73 participants, who were White (47.9%), Asian (23.3%), multiracial (selected more than one race; 21.9%), other (2.7%), African or African American (1.4%), American Indian or Alaskan Native (1.4%) and Pacific Islander (1.4%). We obtained ethnicity information for 77 participants, 17% of whom were Hispanic or Latino, and 83% were not.

Data was collected from February 2022 to July 2022.

Materials

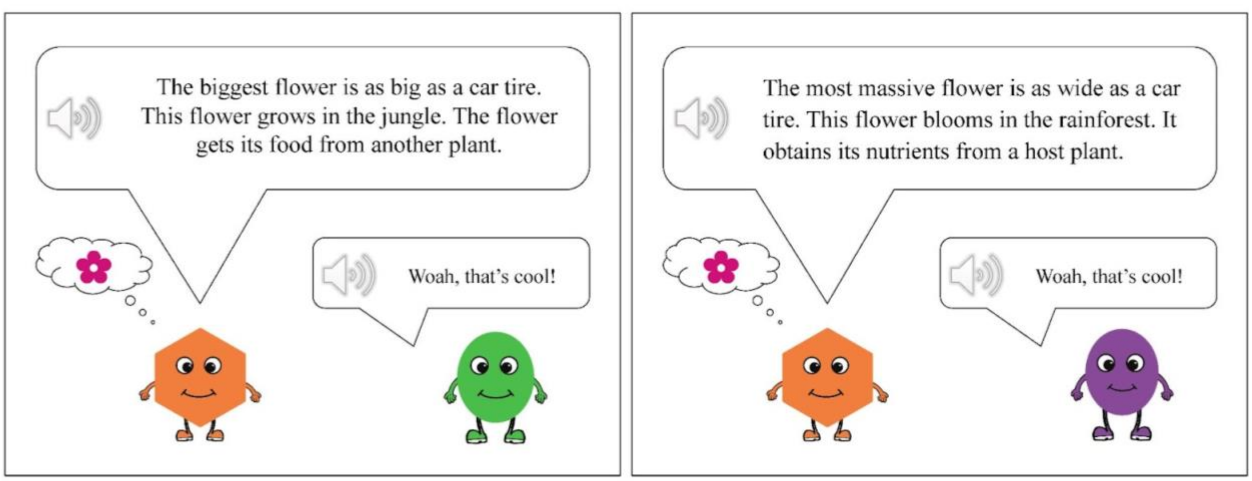
The experiment used an interactive storybook format, presented via PowerPoint. The study plan, hypotheses, and analyses were all preregistered https://aspredicted.org/897_9F6. All study materials are available at https://osf.io/cw8dq/.

Storybook

Participants were introduced to the protagonist, "Figgy," and two fictional groups, "Purpleys" and "Greenies," distinguishable only by color. These groups were introduced as adults who coexisted in Figgy's town. Figgy shared facts with each group through pre-recorded audio clips, consistently employing Low Complexity language for one group and High Complexity language for the other. While the content of the phrases in each pair remained similar, conveying the same facts, the key differences lay in the word choice and delivery. Low Complexity phrases were constructed with simpler words and delivered with exaggerated prosody to mimic child-directed speech. In contrast, High Complexity phrases involved more sophisticated vocabulary and were presented in a formal register with neutral prosody. Participants observed Figgy present six pairs of these phrases, with characters from both groups showing identical levels of understanding. Figure 3.1 provides a trial snapshot. To control for extraneous factors, we counterbalanced the sequence of questions, the order of character presentations, and the initial complexity level of the phrases presented.

Figure 3.1

Schematic of a single trial. In this trial, Figgy addresses a Greenie with Low Complexity language (left panel) and a Purpley with High Complexity language (right panel).



Language Manipulations. We manipulated language complexity in our storybook across three dimensions: age-of-acquisition (AoA), word frequency, and prosody.

Age-of-Acquisition (AoA). AoA reflects the age at which a person remembers learning a particular word (Kuperman et al., 2012). In our study, we employed AoA data from Kuperman et al.'s database, which includes 30,121 English words. To manipulate the complexity of language in our phrases, we replaced simpler words in the Low Complexity phrases with more sophisticated words in the High Complexity phrases. For example, the word “food” in a Low Complexity phrase was substituted with “nutrients” in its High Complexity counterpart. In our AoA analysis, we concentrated on the base forms of these changed words (e.g., using “nutrient” instead of “nutrients”) because the database predominantly includes base words. We specifically analyzed the AoA of the words that were modified in each phrase. This analysis revealed that the words in the High Complexity phrases had a higher average AoA (M = 8.51, SD = 1.99) compared to their counterparts in the Low Complexity phrases (M = 4.33, SD = 1.11). The mean AoA difference for the altered words across trials was 4.18 years ($t(60.46) = 11.69$, $p < .001$), which validates our approach in manipulating language complexity.

Frequency. In linguistics terms, frequency refers to how often words are used within a language context. We selected frequency as a complementary measure to lexical complexity alongside the self-reported AoA data. For this analysis, we utilized the Child Language Data Exchange System (CHILDES), which offers extensive transcripts of child-language interactions (MacWhinney & Snow, 1990). Using the childes-db R API (Sanchez et al., 2019), our analysis focused on the North American English collection, particularly on mother-child conversation transcripts involving children aged 5 to 9 years. We specifically assessed the frequency of the base forms of words that were altered between the Low and High Complexity phrases. This approach ensured consistency with our AoA analysis. Our findings indicated that the altered base words in the High Complexity phrases were used significantly less frequently (M = 592.29 occurrences per 1,000,000 words, SD = 870.73) than those in the Low Complexity phrases (M = 5166.03 occurrences per 1,000,000 words, SD = 5082.82), $t(44.58) = 5.81$, $p < .001$. This frequency analysis complements our AoA results, highlighting that the High Complexity phrases contain words that are not only learned later but also less commonly encountered in the language experiences of children aged 5 to 9 years.

Prosody. We recorded the Low Complexity phrases such that they resembled Child-Directed Speech (CDS), characterized by higher pitch, slower tempo, and exaggerated intonation (Saxton, 2009). High Complexity phrases were recorded with a more formal prosody reflective of adult-directed speech. This modification led to the Low Complexity phrases being spoken slower (M = 148.35 words per minute, SD = 6.85) than the brisker High Complexity phrases (M = 174.17 words per minute, SD = 25.05), $t(5.74) = 2.44$, $p = .05$.

Detailed measures for the altered words, including AoA, frequency, and speed, are documented in Table 3.1.

Table 3.1

Descriptive Statistics for Language Manipulation Measures in Low and High Complexity Phrases

Trial	Language Complexity	Text	Target words	AoA (years)		Frequency (occurrences per 1,000,000 words)		Words per minute
				M	*SD*	*M*	*SD*	
1	Low	The biggest flower is as big as a car tire. This flower grows in the jungle. The flower gets its food from another plant.	biggest, big, grows, jungle, gets, food, another	3.90	1.08	3944.16	1750.17	160
	High	The most massive flower is as wide as a car tire. This flower blooms in the rainforest. It obtains its nutrients from a host plant.	massive, wide, blooms, rainforest, obtains, nutrients, host	8.16	1.61	391.71	559.49	192.31
2	Low	Most living things can see colors. But the colors they see are different. People can see about ten million different colors.	things, see, but, see, people, see, different	3.91	0.98	5322.61	2873.2	140
	High	Most living organisms can detect colors. However, the colors they perceive are different. Humans can detect about ten million distinct colors.	organisms, detect, however, perceive, humans, detect, distinct	8.94	2.17	812.18	728.38	140
3	Low	Jupiter is a planet that goes around the sun. It is the biggest planet. It is two times bigger than all the other planets.	goes, around, biggest, two, times, bigger	4.34	1.63	4183.55	1915.03	143.71
	High	Jupiter is a planet that orbits the sun. It is an enormous planet. It weighs twice as much as all the other planets combined.	orbits, enormous, weighs, twice, much, combined	6.51	1.19	1075.16	1152.79	200
4	Low	When babies are born, they cannot see color. Their vision is in black and white. They start seeing colors when they are five months old.	see, vision, start, seeing	4.27	1.66	5272.03	3868.62	149.7
	High	When babies are born, they cannot perceive color. Their eyesight is in black and white. They begin perceiving colors when they are five months old.	perceive, eyesight, begin, perceiving	8.32	2.67	1825.79	1303.15	186.57
5	Low	People's eyes make tears all the time. Tears are mostly made of water and salt. The salt is made in the body.	people's, make, all, the, time, mostly, made, salt, salt, made	4.56	0.55	6555.66	8959.74	146.66
	High	Human's eyes constantly produce tears. Tears consist primarily of water and sodium. The sodium is produced in the body.	humans, constantly, produce, consist, primarily, sodium, sodium, produced	8.58	1.76	94.95	268.57	146.15
6	Low	Carrots have been grown for thousands of years. But carrots have not always been used just for food. They used to be used as medicine for sickness.	grown, used, just, food, used, to, used, medicine, sickness	4.77	4.77	5058.41	4944.22	150
	High	Carrots have been cultivated for thousands of years. But carrots have not always been utilized exclusively for nourishment. They have historically been utilized as remedy for illness.	cultivated, utilized, exclusively, nourishment, historically, utilized, remedy, illness	9.95	1.55	118.88	336.24	180

Procedure

Warm-up Task

Before the main experiment, we introduced participants to a warm-up task using a character named Bubby. Participants made judgments about Bubby's abilities using thumbs-up/down symbols. This helped familiarize them with the rating scales used later in the study and all participants completed this task successfully.

After the Warm-up Task, participants completed several tasks that aimed to gauge their perceptions of the group members in the story. See Figure 3.2 for a schematic of the various tasks.

Smarter Character Selection

Participants observed Figgy present Low and High Complexity phrases to the Purpleys and Greenies, maintaining consistency in the complexity of language each group received. After each set of phrases, participants judged which character seemed smarter. This task consisted of six trials, with no feedback given.

Group Smartness Ratings

To assess whether participants formed group-level opinions about smartness, we asked them to evaluate both groups' smartness (e.g., "Look at these Greenies. Do you think Greenies are smart or not smart?"). They were prompted to refine their initial binary response to a four-point scale, with 1 signifying "very not smart" and 4 representing "very smart."

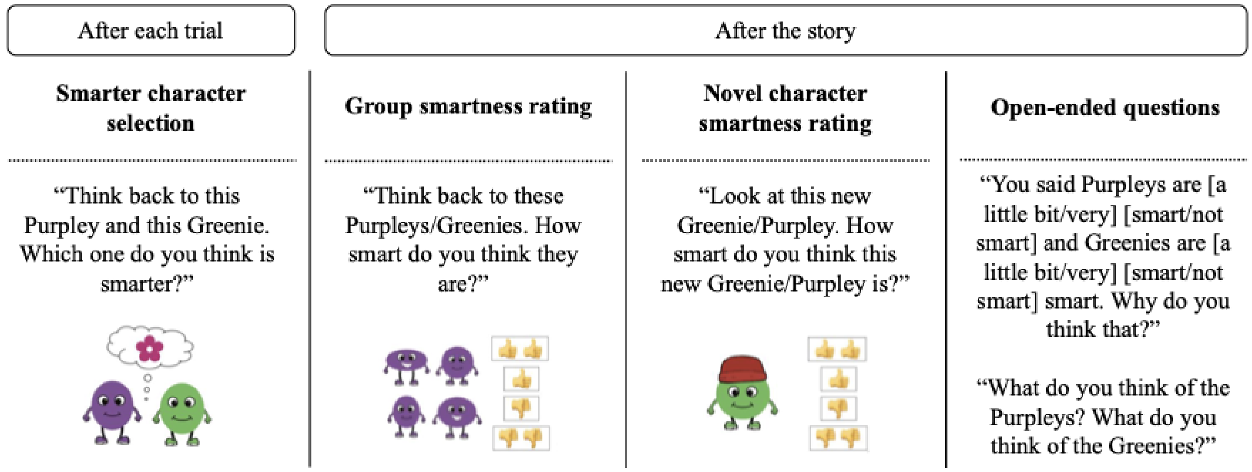
Novel Character Smartness Ratings

Next, we tested whether participants extended their smartness judgments to novel Greenies and Purpleys, that were distinguished from previously seen characters because they were wearing hats. Using the same scale as the previous task, participants evaluated the smartness of the novel characters of each group, and their judgments were recorded on the same 1 ("very not smart") to 4 ("very smart") scale.

Open-Ended Questions

Lastly, we sought qualitative insight into participants' reasoning through open-ended questions about group competence. First, following the group evaluation, participants were reminded of their selection and were asked "Why do you think that?". This question aimed to uncover the underlying factors influencing their judgments. To conclude the experiment, participants were asked what they thought of each group.

Figure 3.2
Schematic of tasks within the study.



Study 1 Results

Smarter Character Selection

We began by determining the percentage of trials in which participants selected the character who heard High Complexity language as being the smarter of the two characters. Five- and six-year-olds' selections did not significantly deviate from chance as they did this in 51% of trials, $SD = 24.17$, $t(31) = 0.12$, $p = .90$). In contrast, eight- and nine-year-olds chose the character who heard the High Complexity phrase in 61% of trials, a rate that is significantly above chance, $SD = 22$, $t(31) = 2.81$, $p = .01$. Adults were even more likely to do so, in 86% of trials, $SD = 17.49$, $t(31) = 11.62$, $p < .001$. See Figure 3.3.

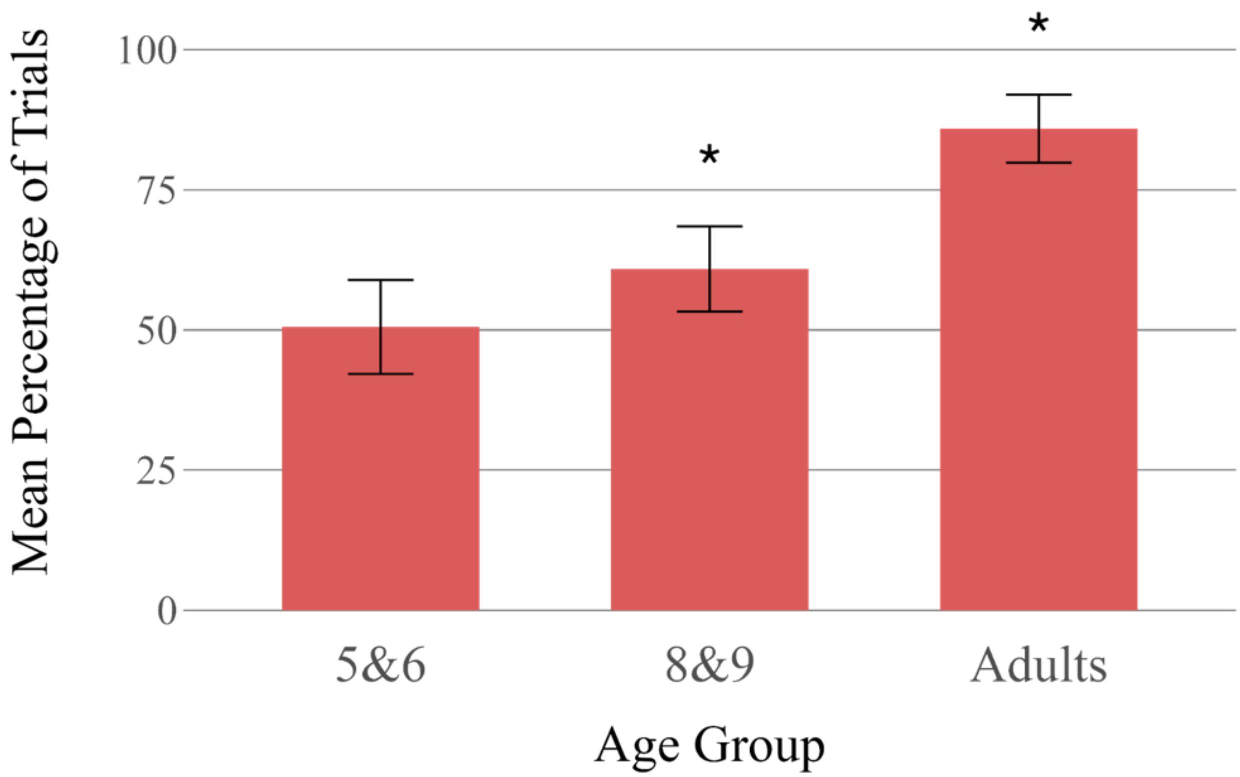
We used a mixed-effects logistic regression model to predict the likelihood of participants choosing the character who heard the High Complexity phrase as smarter in each trial. The model included participants' age group (5-6 years, 8-9 years, or adult), gender (male or female), and trial sequence (1 or 2, indicating whether each trial occurred in the first or second half of the study) as categorical fixed-effect predictors. The trial sequence was included to examine potential learning effects throughout the study. Random intercepts were set for each participant.

A likelihood ratio test comparing our full model (AIC = 679.99) with a null model (AIC = 717.78), which did not consider participant age, indicated a significantly better fit for our model ($\chi^2(2) = 41.79$, $p < .001$). The odds ratio (OR) for 8-9-year-olds choosing High Complexity characters as smarter was 1.59 (95% CI [0.96, 2.63], $p = .07$) compared to 5-6-year-olds. For adults, this tendency was even more pronounced, with an OR of 6.85 (95% CI [3.77, 12.44], $p < .001$), strongly suggesting that adults were more likely to associate High Complexity language with intelligence.

Our analysis also revealed a learning effect across all age groups. Specifically, participants were more likely to choose the character who heard High Complexity language as smarter in the second half of the study (OR = 1.64, 95% CI [1.21, 2.22], $p = .01$). This finding suggests that as participants progressed through the trials, they increasingly associated High Complexity language with intelligence, regardless of their age group.

Figure 3.3

Mean percentage of trials in which participants selected the character who heard High Complexity language as being smarter, by age group. Error bars represent 95% confidence intervals.

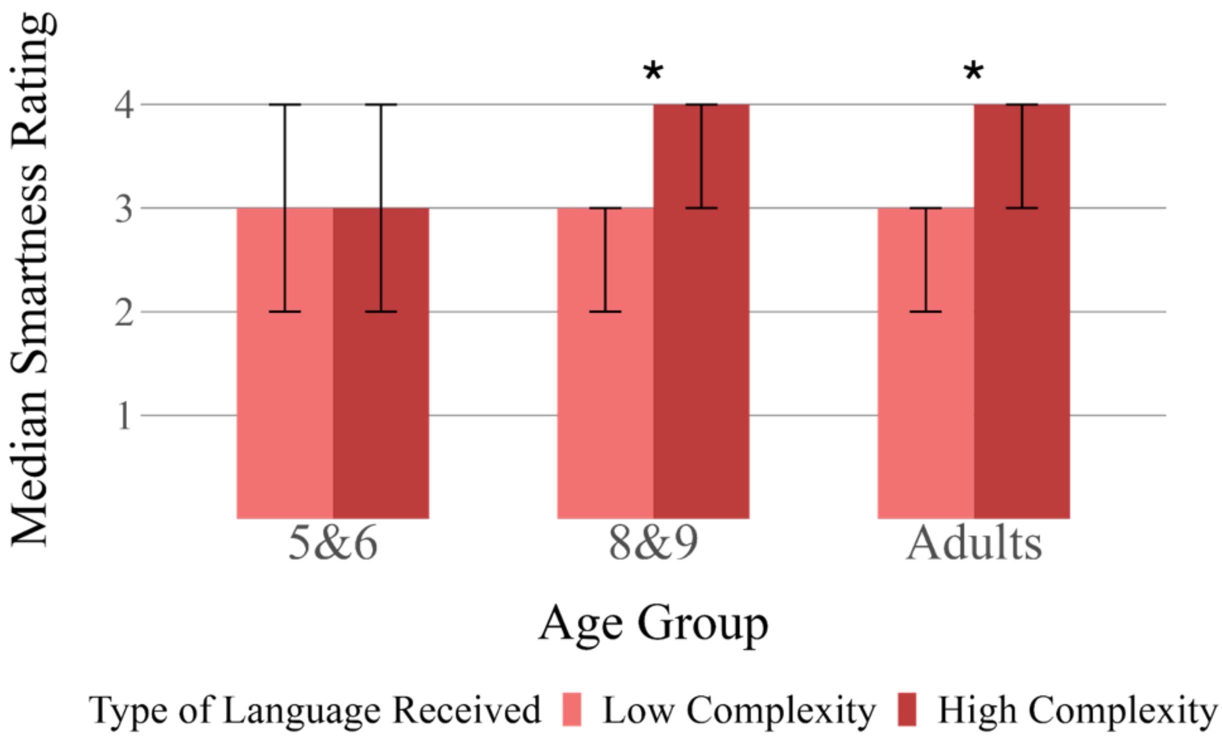


Group Smartness Ratings

Group smartness ratings were recorded on a scale ranging from 1 (“very not smart”) to 4 (“very smart”). Given the non-normality of the distribution, we proceeded with non-parametric tests. We conducted three Wilcoxon signed-rank tests to compare the group ratings in each age group. In the 5-6-year-olds, children’s median ratings for the group addressed with High Complexity language (*Mdn* = 3) did not significantly differ from those addressed with Low Complexity language (*Mdn* = 3), $V = 110$, $p = .60$. However, both 8-9-year-olds and adults rated the group exposed to High Complexity language (*Mdn* = 4) significantly higher than the group exposed to Low Complexity language (*Mdn* = 3), with 8-9-year-olds: $V = 165$, $p < .001$, and adults: $V = 259$, $p = .001$. See Figure 3.4.

Figure 3.4

Median group smartness ratings by age group and type of language complexity received. High Complexity language included rarer, later-acquired words and was spoken with adult-directed prosody. Error bars represent the interquartile range.

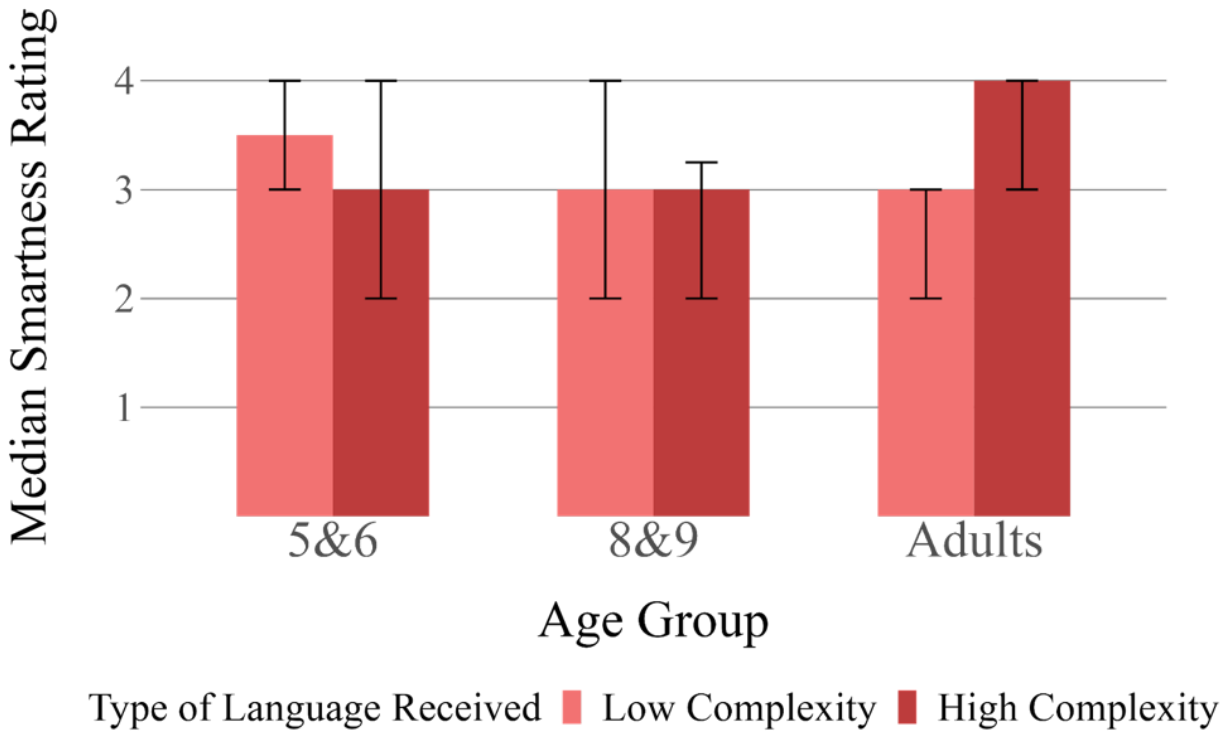


Novel Character Smartness Ratings

Group smartness ratings were recorded on the same scale ranging from 1 (“very not smart”) to 4 (“very smart”). Given the non-normality of the distribution, we again proceeded with non-parametric tests. We conducted three Wilcoxon signed-rank tests to compare the novel character ratings in each age group. For participants aged 5-6 years, no significant difference emerged between their smartness ratings for the novel character belonging to the group who received High Complexity language (*Mdn* = 3) and the character belonging to the group who received Low Complexity language (*Mdn* = 3.5), ($V = 68$, $p = 0.28$). This trend was mirrored in the 8-9-year-old participants, who rated both novel characters as a median of 3, $V = 40$, $p = 0.25$. Adults, however, rated the novel character who belonged to the group who received High Complexity language (*Mdn* = 4) as significantly smarter than the character belonging to the group who received Low Complexity language (*Mdn* = 3), ($V = 171$, $p < .001$). See Figure 3.5.

Figure 3.5

Median smartness ratings for novel characters by language complexity group. Novel characters were associated with either the group that received Low Complexity language or the group that received High Complexity language. Error bars represent the interquartile range.



Open-Ended Questions

We wanted to explore whether explicit knowledge of the variation in Figgy's language complexity shaped participants' responses in the forced-choice task. To address this, we first identified participants who explicitly mentioned Figgy's manner of talking, including comments about vocabulary richness, speech tempo, and other relevant linguistic characteristics, in any of the open-ended questions. Two coders independently coded this, yielding a Cohen's Kappa inter-rater reliability score of 0.75. This score, falling between the 0.61 to 0.89 confidence interval range, indicates substantial agreement between coders. Any coding discrepancies were resolved by a third, independent coder. One 5-6-year-old (3%), 8 8-9-year-olds (25%), and 25 adults (78%) mentioned language.

We next ran an exploratory model which included all variables in the full model (age group, gender, trial sequence, and a random intercept for participants), and whether participants mentioned language in an open-ended response as categorical fixed-effect predictors, with random intercepts for participant. In this model, explicitly mentioning language significantly increased participants odds of selecting the character who heard the High Complexity phrase as being smarter in the forced-choice task (OR = 3.74, 95% CI = [2.02, 6.90], $p < .001$). A likelihood ratio test suggests that this full model with open-ended responses (AIC = 664.47) resulted in a significantly better fit than the full model (AIC = 679.99), which did not include open-ended responses as a predictor ($\chi^2(2) = 17.52$, $p < .001$). This suggests that explicit awareness of variation in language complexity may play a key role in this inferential process.

Study 1 Discussion

Study 1 investigated how observers—children and adults—judge an individual's competence based on the complexity of language directed at them. Characters in a storybook were addressed with either Low or High Complexity language, differing in lexical sophistication and prosody.

Our results reveal a developmental trend: while younger children did not significantly differentiate between the two language complexities, older children and adults were more inclined to perceive characters addressed with High Complexity language as more competent. This pattern suggests that the tendency to infer competence from observed language interactions develops with age.

Group smartness ratings varied across age groups, highlighting that the ability to form group-level competence judgments based on language likely requires cognitive skills that evolve over time. This developmental progression was especially evident in adults, who not only discerned competence differences based on language complexity but also extended these judgments to novel characters associated with the High Complexity language group. Such a trend suggests a maturation in cognitive abilities, enabling adults to generalize competence-related stereotypes from specific instances to broader group contexts. In contrast, younger children displayed no clear preference in their competence perceptions, reflecting their still-developing capacity to abstract and apply these linguistic observations across different individuals and contexts.

The varied responses to open-ended questions shed light on the developmental differences in perceiving and interpreting variations in a speaker’s language complexity. Adults, who most frequently mentioned aspects like vocabulary richness and prosody, were significantly more likely to associate characters addressed with High Complexity language as being smarter. This suggests that explicit awareness of linguistic nuances may be a critical factor in this inferential process. Such awareness, correlating with age, underscores the intricate interplay between cognitive development and social perception.

Study 2: Impact of Content Complexity on Perceptions of Competence

Building on Study 1's findings that linguistic complexity influences perceptions of competence, Study 2 explores the role of content complexity. This aspect of communication, exemplified in scenarios like “mansplaining,” where not only language but also content is simplified, may similarly impact observers' competence inferences.

In Study 2, we maintain a similar setup to Study 1, where observers—children and adults—witness an individual being addressed with either Low or High complexity content. This time, instead of focusing on the linguistic form, we shift our attention to the content of the communication. We specifically contrast mundane, well-known facts (Low Complexity) with unusual, intriguing information (High Complexity) to see if and how this affects the observers' judgments of the addressed individual's competence. The aim is to determine whether the observers, particularly younger children, are more attuned to content complexity as a cue for inferring competence, providing further insights into their cognitive processing of communicative nuances.

Study 2 also provides insight into why younger children in Study 1 might not have associated linguistic complexity with competence. Were their perceptions unaltered due to an overall insensitivity to linguistic variations, or due to a cognitive limitation in processing abstract linguistic cues? Understanding these mechanisms is pivotal in furthering our comprehension of how children and adults discern competence through communication.

Method

Participants

We tested 96 new participants. The sample again consisted of 32 5-6-year-old children (*M* = 5.84 years, *SD* = 0.61, 16 females, 16 males), 32 8-9-year-old-children (*M* = 8.95 years, *SD* = 0.58, 16 females, 16 males), and 32 adults (*M* = 21.69 years, *SD* = 1.52 years, 16 females, 16 males).

We obtained race information for 68 participants. Participants were White (49%), Asian (26%), multiracial (selected more than one race; 21%), other (3%), and African or African American (1%). We obtained ethnicity information for 75 participants. Sixteen percent of participants were Hispanic or Latino, and 84% were not.

Data was collected from November 2022 to September 2023.

Materials

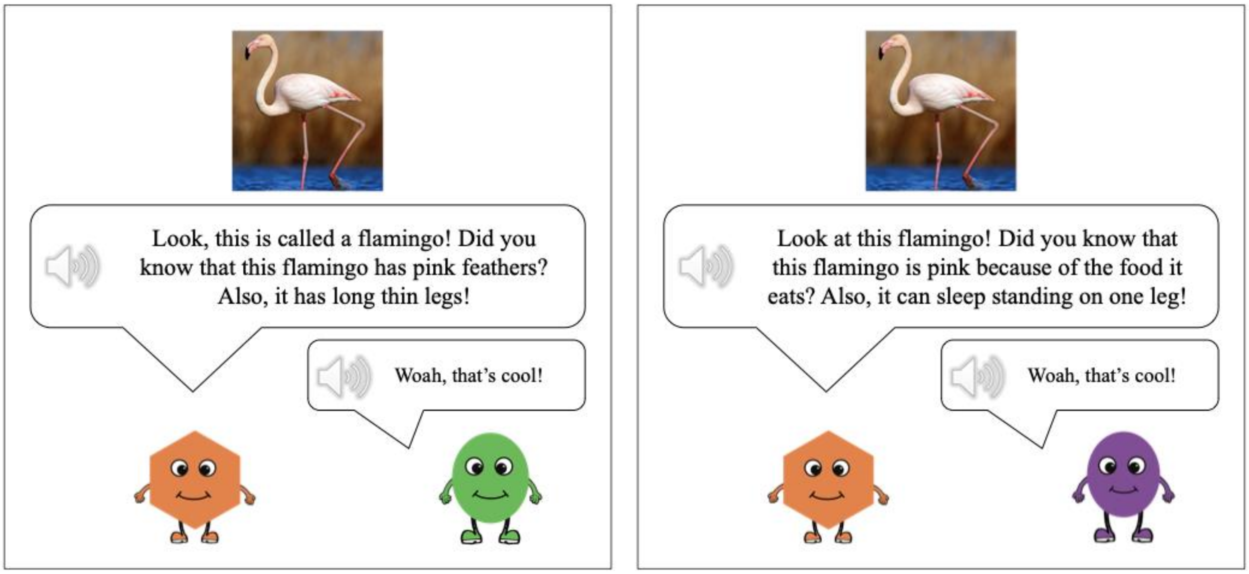
Similar to Study 1, participants interacted with an experimenter through a storybook presented via PowerPoint. To accommodate the shift in focus from linguistic complexity to content complexity, we made specific modifications to the storybook while maintaining the basic structure for consistency. The study plan, hypotheses, and analyses were all preregistered at https://aspredicted.org/6MJ_XJZ.

Storybook

The storybook for Study 2 maintained the character of Figgy and the two fictional groups of Purpleys and Greenies. Key changes were made to shift the focus from linguistic complexity to the content of the information being presented. Figgy again communicated facts to the Purpleys and Greenies, but instead of varying linguistic complexity, Figgy presented information that varied in depth and familiarity. Low Complexity information was common knowledge and not interesting, whereas High Complexity information was less commonly known and more intriguing. The topics were identical across the Low Complexity and High Complexity phrases, and they were similar in linguistic complexity and length. Thus, this manipulation allowed us to tease apart the influence of content complexity in shaping perceptions of competence, separate from the linguistic factors explored in Study 1. See Figure 3.6 for a trial snapshot.

Figure 3.6

Schematic of a single trial. In this trial, Figgy addresses a Greenie with Low Complexity information (left panel) and a Purpley with High Complexity information (right panel).



Content Manipulations

To validate the complexity of the information presented in the storybook, we conducted a preliminary assessment with 20 adult participants recruited from Prolific. This assessment focused on two key aspects: the interestingness and commonality of each fact presented in the storybook.

Participants rated the interestingness of each fact on a scale from 0 (not at all interesting) to 100 (extremely interesting). They also estimated the familiarity of each fact, using a scale from 0% (unknown to most) to 100% (known by all). Using these ratings, we calculated a "complexity score" for each fact by reverse-coding the familiarity ratings and then averaging these with the interestingness scores. Higher scores were assigned to facts that were less commonly known and found more interesting.

From this assessment, we selected six pairs of facts for the storybook. Each pair comprised one Low Complexity fact (with lower complexity scores) and one High Complexity fact (with higher complexity scores). The High Complexity facts had an average interestingness score of 56.98, while the Low Complexity facts averaged 10.03 on the same scale. This disparity in scores confirmed the effectiveness of our manipulation in differentiating the complexity of the facts.

Procedure

We employed measures similar to those in Study 1 to evaluate participants' perceptions based on the complexity of content addressed to characters in the storybook.

Warm-Up Task

Before the main experiment, participants completed a warm-up task–identical to that used in Study 1–to familiarize themselves with the rating scales used in the study. All participants completed this task successfully.

Smarter Character Selection

In the storybook, Figgy relayed various facts to each group, consistently sharing facts that had Low Complexity information with one group, and High Complexity information with the other. Following each set of facts, participants were asked to select which character they think is more intelligent, without receiving any feedback.

Group Smartness Ratings

Following the same procedure as Study 1, we assessed whether participants had formed group-level beliefs about the Greenies' and Purpleys' competence. Responses were recorded on a four-point scale, with '1' signifying "very not smart" and '4' representing "very smart."

Novel Character Smartness Ratings

Similar to Study 1, participants were asked to rate the smartness of novel characters associated with each group. These characters, distinguished by wearing hats, were rated using the same four-point scale, allowing us to see if perceptions of intelligence were generalized to new members of each group.

Generalized Competence Assessments

This study included an additional assessment to address a potential confound. Specifically, we aimed to determine if the group exposed to High Complexity information was considered "smarter" solely due to their recent exposure to novel facts, or if there were deeper perceptions of inherent competence. To this end, participants were shown images of both the Greenies and Purpleys and were asked to assess which group was better at "solving hard puzzles" and which group "learns things really quickly." These competencies, known to be equated with intelligence in prior research with children (Baharloo et al., 2022; Bian et al., 2017), were chosen as they reflect innate intellectual abilities, distinct from knowledge gained from recent interactions with Figgy. This approach helped us to investigate if participants' judgments of intelligence were based on intrinsic capabilities of the characters, rather than just the immediate facts they received.

Open-Ended Questions

As in Study 1, participants in Study 2 responded to open-ended questions about the groups' competence. After their group evaluations, we asked, "Why do you think that?" to explore the reasoning behind their judgments. Participants were also asked to share their general impressions of each group. This method was identical to Study 1, aiming to gain qualitative insights into participants' cognitive processes.

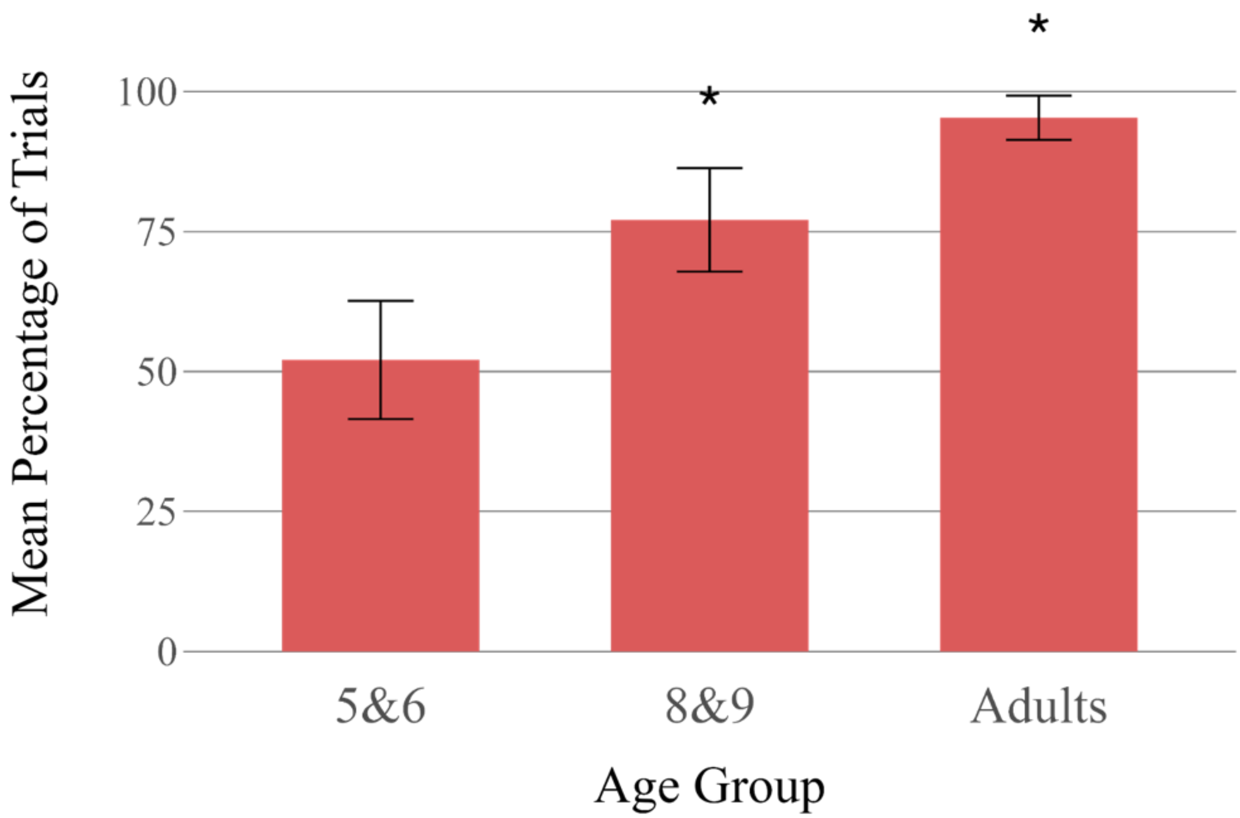
Study 2 Results

Smarter Character Selection

As with Study 1, we first assessed the percentage of trials in which participants identified the character who heard High Complexity content as smarter. Five- and six-year-olds perceived these characters as smarter in 52% of trials, not differing significantly from chance, $SD = 30.49$, $t(31) = 0.37$, $p = .71$. In contrast, both 8-9-year-olds and adults showed a clear trend towards selecting the High Complexity character as smarter, with 8-9-year-olds doing so in 77% of trials, $SD = 26.71$, $t(31) = 5.72$, $p < .001$, and adults in 95% of trials, $SD = 11.41$, $t(31) = 22.46$, $p < .001$. See Figure 3.7.

Figure 3.7

Mean percentage of trials in which participants selected the character who heard High Complexity content as being smarter, by age group. Error bars represent 95% confidence intervals.



We employed a mixed-effects logistic regression model to predict the likelihood of participants choosing the character who heard High Complexity content as smarter. The model included participants' age group (5-6 years, 8-9 years, or adult), gender (male or female), and trial sequence (1 or 2, indicating whether each trial occurred in the first or second half of the study) as

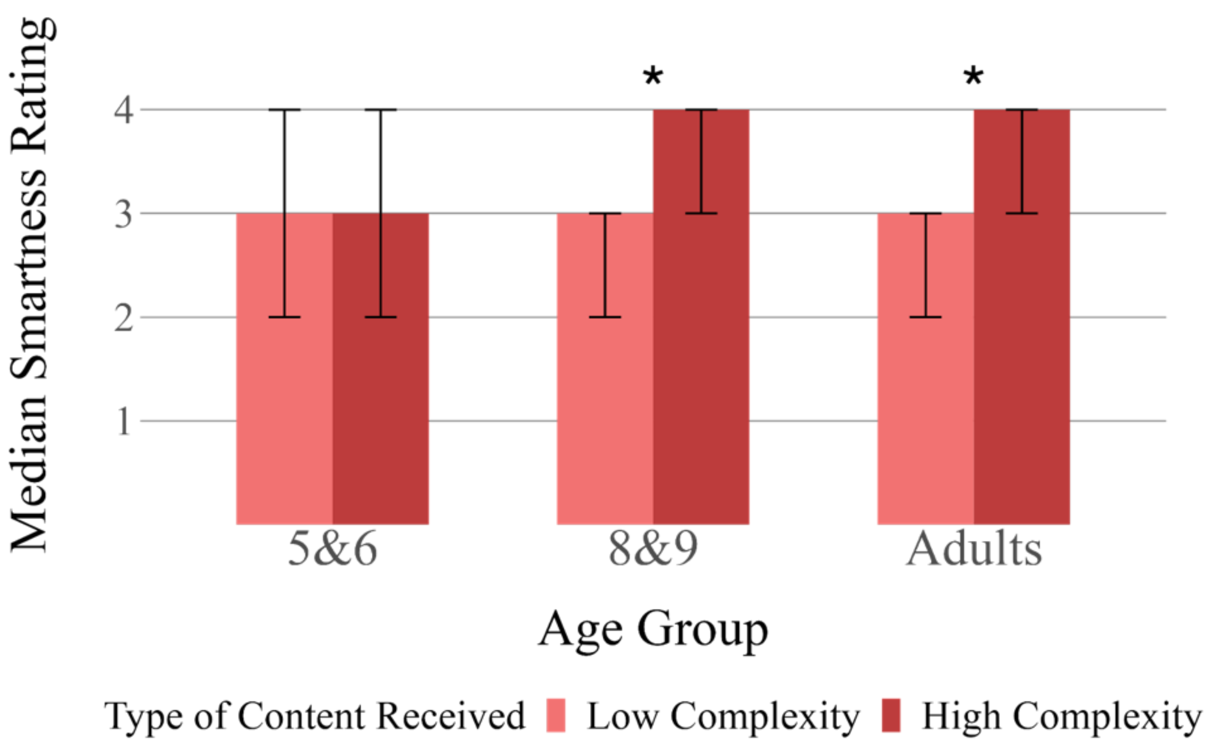
categorical fixed-effect predictors. The trial sequence was included to examine potential learning effects throughout the study. Random intercepts were set for each participant.
Our full model (AIC = 515.57) showed a significantly better fit than the null model (AIC = 557.52), which excluded participant age ($\chi^2(2) = 45.952$, $p < .001$). The odds ratio for 8-9-year-olds was 4.82 (95% CI [1.92, 12.04], $p < .001$), and for adults, it was 43.44 (95% CI [13.00, 145.13], $p < .001$), indicating a strong age-related trend in associating High Complexity content with intelligence. Unlike Study 1, no learning effect was observed in trial sequence (OR = 1.09, 95% CI [0.68, 1.76], $p = .72$), suggesting consistent responses throughout the study.

Group Smartness Ratings

Similar to Study 1, we used non-parametric tests for group smartness ratings, given the non-normal distribution of responses. In this measure, 5-6-year-olds showed no significant difference in their ratings between the group that heard High Complexity content (*Mdn* = 3) and the group that heard Low Complexity content (*Mdn* = 3), $V = 190$, $p = .99$. This suggests that younger children did not distinguish between the groups in terms of perceived intelligence. In contrast, both 8-9-year-olds and adults rated the group that heard High Complexity content as significantly smarter (*Mdn* = 4) compared to the group that heard Low Complexity content (*Mdn* = 3). This was evident from the Wilcoxon signed-rank test results for 8-9-year-olds ($V = 346.5$, p $< .001$) and adults ($V = 231$, $p < .001$), indicating a clear developmental trend in associating content complexity with intelligence. See Figure 3.8.

Figure 3.8

Median group smartness ratings by age group and content complexity type. Error bars represent the interquartile range.

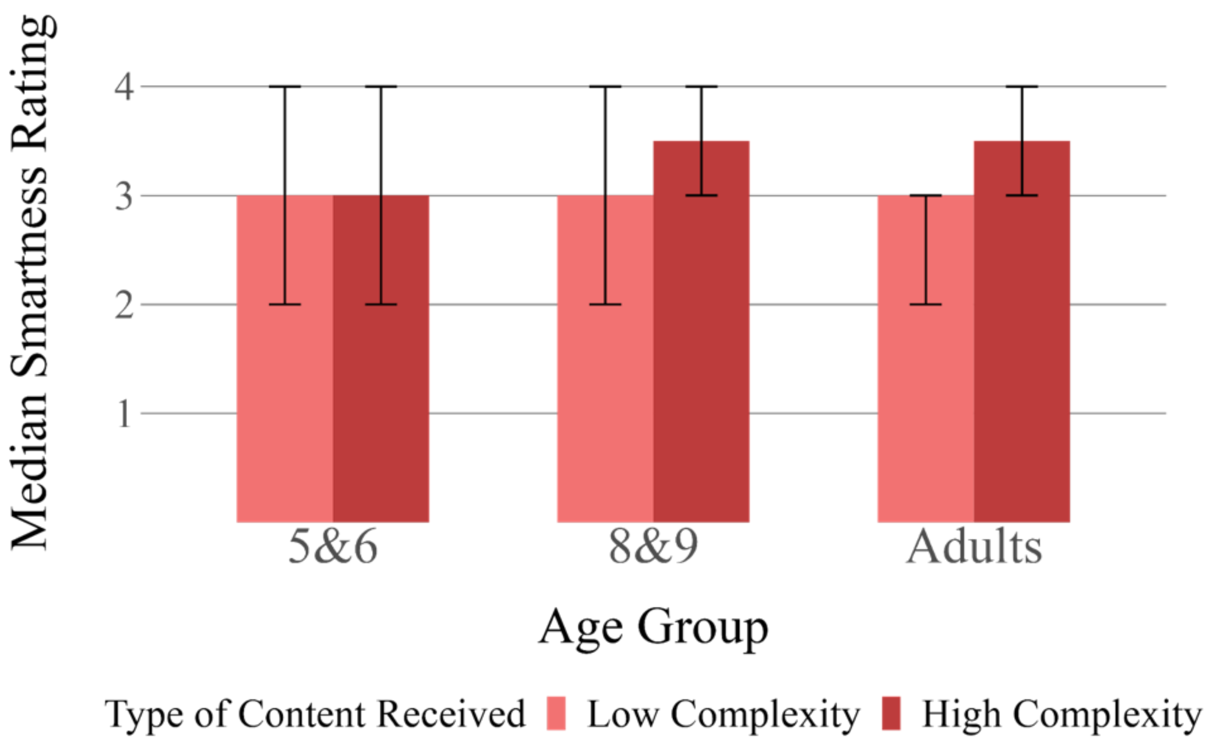


Novel Character Smartness Ratings

Non-parametric Wilcoxon signed-rank tests, used due to the non-normal distribution of ratings, revealed distinct developmental patterns in how novel characters were perceived based on content complexity. Children aged 5-6 and 8-9 years showed no significant difference in their ratings for novel characters associated with High Complexity (5-6 years: *Mdn* = 3; 8-9 years: *Mdn* = 3.5) versus Low Complexity (5-6 years: *Mdn* = 3; 8-9 years: *Mdn* = 3) groups. These results suggest that children across these age groups did not extend their judgments of intelligence based on content complexity to new characters (5-6 years: $V = 107.5$, $p = .35$; 8-9 years: $V = 147$, $p = .11$). In contrast, adults demonstrated a marked tendency to attribute higher intelligence to novel characters from the High Complexity group (*Mdn* = 4) compared to the Low Complexity group (*Mdn* = 3), $V = 231$, $p < .001$. This finding indicates a developmental progression in adults' ability to generalize competence attributions based on the complexity of content. See Figure 3.9.

Figure 3.9

Median smartness ratings for novel characters by content complexity group. Novel characters were associated with either the group that received Low Complexity content or the group that received High Complexity content. Error bars represent the interquartile range.



Generalized Competence Assessments

To determine if participants' judgments of intelligence were influenced solely by the complexity of information recently acquired, rather than inherent competence, we assessed their perceptions of each group's ability to solve puzzles and learn quickly. Results showed that the five- and six-year-olds, selected the High Complexity group at chance, in 52% of trials, indicating their judgments were not swayed by the complexity of information (*SD* = 26.89, $t(31) = 0.33$, $p = .75$). In contrast, eight- and nine-year-olds favored the High Complexity group (61% of trials, *SD* = 27.63, $t(31) = 2.24$, $p = .03$). Among adults, the preference for the High Complexity group was more pronounced (83% of trials, *SD* = 24.13, $t(31) = 7.69$, $p < .001$), indicating that they were more likely to infer inherent intelligence based on the complexity of the information presented. These results strongly suggests that for older children and adults, the perception of competence extended beyond the immediate context of the storybook, countering the potential confound that judgments were based solely on recent exposure to novel information.

Open-Ended Responses

We analyzed participants' open-ended responses to understand if their explicit recognition of the complexity of information influenced their forced-choice decisions. Participants' references to the novelty, depth, or intricacy of Figgy's statements were independently coded, achieving a Cohen's Kappa inter-rater reliability of 0.64. This score, falling between the 0.48 to 0.80 confidence interval range, indicates substantial agreement between coders. Any coding discrepancies were resolved by a third, independent coder. Only a small proportion of the youngest group (5-6-year-olds, 3%) mentioned the quality of information. In contrast, a substantial number of 8-9-year-olds (38%) and a majority of adults (66%) recognized and mentioned the complexity of Figgy's statements.

We constructed an exploratory model that incorporated age group, gender, trial sequence, and explicit mention of information complexity in responses, as well as a random intercept for participants, as predictors. This model demonstrated a significant relationship between recognizing information complexity and perceiving characters who heard High Complexity information as smarter (OR = 5.17, 95% CI [1.69, 15.82], $p = 0.004$). The model with open-ended responses (AIC = 508.64) showed a significantly better fit than the model excluding these responses (AIC = 515.57), $\chi^2(2) = 8.93$, $p = 0.003$. These results suggest that explicit awareness of the depth and novelty of information plays a significant role in how participants, especially older children and adults, infer intelligence based on the content presented to characters.

Study 2 Discussion

Building on the findings of Study 1, which highlighted the impact of linguistic complexity on perceiving intelligence, Study 2 shifted focus to content complexity. This exploration provided further insights into how different aspects of language can shape competence stereotypes.

The results of Study 2 largely paralleled those from Study 1. In both studies, participants, particularly older ones, were more inclined to perceive characters exposed to High Complexity speech (whether in terms of language or information) as more intelligent. This consistency was evident in the forced-choice task, group smartness ratings, and novel character assessments.

These findings support our hypothesis that the ability to form and generalize competence judgments based on linguistic cues relies on cognitive skills that mature over childhood.

Study 2 underscored that not only the form but also the content of language significantly influences perceptions of competence. Characters presented with interesting and less-known High Complexity information were perceived as more competent than those who received mundane, well-known Low Complexity facts. However, the specific reasons for younger children's inability to make similar inferences remain ambiguous. Two possibilities emerge: either they are unable to discern the differences between Low and High Complexity content, or they fail to associate these differences with group membership and perceived competence. Study 3 aims to investigate the former hypothesis, shedding light on the cognitive processes underlying these observations in young children.

Study 3: Assessing Young Children's Ability to Discern Content Complexity

In Studies 1 and 2, we observed that unlike older children and adults, younger children did not infer competence from the complexity of Figgy's speech. This observation raises a critical question: Does this difference arise from younger children's inability to distinguish between Low and High Complexity speech, or does it reflect broader limitations in their inferential reasoning? Study 3 seeks to investigate this distinction. Specifically, our aim is to determine whether younger children can differentiate between the Low and High complexity content presented in Study 2.

Method

Participants

A total of 32 new participants, ages 5 and 6 (M = 5.87, SD = 0.57, 16 females, 16 males), participated in the study.

We obtained race information for 18 participants. Participants were White (45%), multiracial (selected more than one race; 33%), and Asian (22%). We obtained ethnicity information for 17 participants. Twelve percent of participants were Hispanic or Latino, and 88% were not.

Data was collected from October 2023 to November 2023.

Materials

The study plan, hypotheses, and analyses were all preregistered at https://aspredicted.org/JK6_HLK.

Storybook

Participants engaged with a storybook presented through Qualtrics, featuring several pages, each displaying two side-by-side cards. While the cards shared identical images, they differed in content: one card contained Low Complexity (i.e., mundane, and well-known) information related to the image, and the other, High Complexity (i.e., interesting and uncommon) information related to the image. These facts were identical to the ones presented in Study 2. Each participant saw six pairs of cards. We counterbalanced question sequence and randomized the presentation order of the High Complexity and Low Complexity cards.

Procedure

Familiarization Task

Initially, children were introduced to the concepts of "interesting" and "boring" through definitions and examples. They then viewed two card pairs: one boring (Low Complexity) and one interesting (High Complexity) per pair. The task required them to identify the more interesting card in each pair. Correct responses were affirmed; incorrect answers received gentle correction with appropriate clarifications.

More Interesting Content Selection

After the familiarization task, participants were presented with 6 more pairs of cards, each containing a Low Complexity and a High Complexity card. After reading both cards in the pair, the experimenter asked the child to indicate which card they thought was more interesting. Unlike the familiarization task, the experimenter did not provide any feedback. Across all trials, we coded whether children selected the Low Complexity or High Complexity card.

Study 3 Results

We started our analysis by calculating the percentage of trials in which participants correctly selected the interesting card. Children succeeded in this task in 80.73% of trials, significantly above chance (SD = 23.61, $t(31) = 7.36$, $p < .001$). This finding indicates that children were generally adept at distinguishing between more and less interesting content.

We then used a generalized mixed-effects logistic regression model to predict the likelihood of participants accurately choosing the more interesting card. This model considered participant age as a continuous variable, gender, and trial sequence (first half or second half of the study) as fixed-effect predictors, with random intercepts for each participant. The analysis showed a significant age effect (OR = 6.39, 95%CI [1.85, 22.11], $p = .003$), suggesting an increase in the odds of choosing correctly with each additional year of age. This (AIC = 174.49) fitted the data significantly better than a null model without age as a predictor (AIC = 181.81, $\chi^2(2) = 9.33$, $p = 0.002$).

We next explored children's performance on the familiarization task. Children, on average, answered the familiarization task questions correctly 76.6% of the time, SD = 30.7, $t(31) = 4.84$, $p < .001$, a rate significantly higher than chance. This finding suggests that most children had a good initial understanding of what constituted "interesting" versus "boring" content. Subsequently, we conducted exploratory analyses to assess the influence of children's performance on the familiarization task on their subsequent choices in the test trials. To this end, we constructed a third logistic regression model that included all predictors from the full model, with the addition of the number of trials answered correctly in the familiarization task. A likelihood ratio test comparing this third model, including familiarization task performance (AIC

= 169.92), with the previous full model (AIC = 174.49), indicated a significantly better fit ($\chi^2(2)$ = 6.5647, p = 0.010). This model indicated that the number of familiarization tasks answered correctly was a significant predictor for selecting the interesting card in test trials (OR = 3.00, 95% CI [1.28, 7.03], p = 0.011). This result implies that children who performed well in the familiarization task were more likely to consistently identify the interesting card in the test trials.

Study 3 Discussion

Study 3 investigated younger children's ability to distinguish between mundane, Low Complexity content and interesting, High Complexity content. This investigation was essential to understand their competence inferences observed in Studies 1 and 2. The results demonstrated that most 5 to 6-year-olds could reliably differentiate between Low and High Complexity information, indicating that their limited competence inferences in the earlier studies were not due to an inability to perceive content complexity.

The study's findings also revealed a developmental trend: children's capability to discern interesting content improved with age, suggesting a growth in cognitive skills related to content evaluation. The correlation between performance in the familiarization task and the test trials underscores the importance of early understanding of content quality in children's ability to discern.

These findings suggest that younger children’s performance in Studies 1 and 2 may lie more in linking their perceptions of speech complexity to broader social judgments, rather than in the perception of complexity itself. This insight is critical for understanding the development of competence stereotypes and the complex cognitive skills involved in this process.

Discussion

Our series of studies investigated how the complexity of a speaker's communication influences observers' perceptions of a listener's competence across different developmental stages. In Study 1, we found that older children and adults infer higher competence when individuals are addressed with High Complexity language, characterized by advanced lexical choice and prosody. Study 2 extended this finding to the domain of content complexity, revealing that both children and older adults perceive individuals as more competent when they are presented with novel and interesting information, as opposed to mundane information. Study 3 focused on younger children (5-6 years) and demonstrated that they can differentiate between high and low complexity content, suggesting that their performance in Studies 1 and 2 is not due to an inability to differentiate content complexity. Together, these studies highlight a developmental trajectory in the use of linguistic cues as markers of social judgment and competence.

Developmental Trajectory in Sensitivity to Linguistic Cues
Across Studies 1 and 2, we observed a developmental trend where individuals increasingly relied on language complexity as a criterion for assessing listeners’ competence. This trend was particularly pronounced in older children and adults, who consistently linked higher competence with characters addressed with High Complexity language or content. This finding suggests a

developmental shift in recognizing and using linguistic cues as markers of social judgment. However, younger children's lack of association with these cues highlights a potential developmental threshold in their ability to apply linguistic observations in forming competence judgments.

Influence of Language on Stereotype Formation and Reinforcement

Our findings align with existing literature suggesting that language is a powerful tool in shaping social perceptions, including stereotypes. This series of studies underscores the impact of both language complexity and content complexity in reinforcing or possibly creating competence-related stereotypes. Significantly, our research further highlights the power of indirect linguistic cues. Neither the High nor Low Complexity speech explicitly belittled or praised the listener. Yet, the indirect message sent by the type of language used was clear to older children and adults, who inferred a level of competence based on these subtle cues. This complements Grice's Cooperative Principle (1975), highlighting how much communication relies on implicit understandings and context, contributing to the nuanced process of stereotype reinforcement.

Educational Implications

The insights from our studies have considerable implications for educational settings. Given that even subtle linguistic differences can influence children's and adults' perceptions of competence, educators and caregivers should be mindful of the complexity of the language they use when addressing individuals, particularly in group settings. Considering classroom dynamics, these findings become even more salient. Previous studies, like those by Mckown et al., (2010) show how deeply stereotypes can influence educational interactions. Children are sensitive to subtle behaviors that reflect teachers' expectations of students (e.g., the amount of negative feedback a teacher provides and the extent to which teachers require students to adhere to rules). This, in turn, often results in stereotype threat, in which a student inadvertently fulfills a teacher's low expectations. Our data suggest a novel contribution to this work: variations in communication style can transmit expectations about competence or performance, potentially shaping students' trajectories.

Implications for Confirmation of Stereotypes

Stereotype confirmation is not only an issue in classrooms. Listeners of all ages may unknowingly affirm speakers' stereotypes through linguistic entrainment, responding to simplified speech in a way that is similarly matched in complexity (Abney et al., 2014). This response not only validates the speaker's initial stereotype but also potentially shapes onlookers' perceptions. It is imperative for future research to delve into linguistic simplification's real-world effects, particularly on listeners' behaviors and emotions.

Conclusion

In conclusion, our series of studies highlight the profound impact of communication complexity on shaping perceptions of competence. Starting from a young age, individuals are sensitive to the ways in which speakers adjust their language when addressing different social groups. This sensitivity evolves with age, influencing how individuals use linguistic cues to form judgments about others' competence. Future research should aim to further unravel the underlying mechanisms behind these observations and develop strategies to mitigate the unintentional transmission of stereotypes through language.

Conclusion

This dissertation explores the development of stereotypes and highlights that stereotypes are not only internalized from an early age but are also perpetuated through nuanced linguistic cues and social interactions. The research presented offers new insights into the mechanisms underlying stereotype transmission and suggests potential avenues for intervention.

Chapter 1 unveils that children as young as 7 internalize racial stereotypes akin to those prevalent among U.S. adults, challenging assumptions about the roots of these biases. This finding underscores the urgency of early interventions to dismantle stereotypes and raises a pivotal question that informs the subsequent chapters: How do children learn these specific stereotypes?

Chapter 2 uncovers the subtle yet profound role of linguistic cues in stereotype formation. The evidence in this chapter suggests that children and adults infer judgments about groups based on indirect linguistic cues, highlighting the necessity for mindful and inclusive language in varied settings like education and media.

Chapter 3 expands this work by demonstrating how observers' perceptions of competence are influenced by the manner in which individuals are addressed. This underscores the impact of first-person linguistic interactions on stereotype formation and reinforcement.

This research lays the foundation for various future endeavors. In educational contexts, developing interventions that encourage teachers to adopt uniform communication styles with all students could be a significant step. In the realm of media, analyzing how character portrayals and dialogues reinforce or challenge stereotypes can guide more responsible content creation. Additionally, it will be critical to examine the psychological and behavioral impact of differential linguistic treatment on individuals facing bias and prejudice, as this could unintentionally lead to the confirmation of stereotypes (e.g., through stereotype threat, linguistic entrainment, etc.).

In conclusion, this dissertation contributes to a deeper understanding of the origins and development of stereotypes. It sheds light on how seemingly innocuous aspects of daily communication can reinforce societal biases, and highlights the critical role of direct linguistic interactions, an aspect previously unexplored. This work also paves the way for tangible societal change. I am hopeful that these insights will inspire a shift in how we communicate about and to diverse social groups, fostering a more equitable society.

References

Abney, D. H., Paxton, A., Dale, R., & Kello, C. T. (2014). Complexity matching in dyadic conversation. *Journal of Experimental Psychology: General*, *143*(6), 2304–2315. https://doi.org/10.1037/xge0000021

Aboud, F. E. (2003). The formation of in-group favoritism and out-group prejudice in young children: Are they distinct attitudes? *Developmental Psychology*, *39*(1), 48–60. https://doi.org/10.1037/0012-1649.39.1.48

Ahl, R. E., & Dunham, Y. (2019). "Wealth Makes Many Friends": Children Expect More Giving From Resource-Rich Than Resource-Poor Individuals. *Child Development*, *90*(2), 524–543. https://doi.org/10.1111/cdev.12922

Asaba, M., Wu, Y., Carrillo, B., & Gweon, H. (2020). You're surprised at her success? Inferring competence from emotional responses to performance outcomes. *Proceedings of the 42nd Annual Conference of the Cognitive Science Society*, 2650–2656.

Baharloo, R., Vasil, N., Ellwood-Lowe, M. E., & Srinivasan, M. (2023). Children's use of pragmatic inference to learn about the social world. *Developmental Science*, *26*(3), e13333. https://doi.org/10.1111/desc.13333

Baharloo, R., Xu, F., & Bian, L. (2022). *The development of racial stereotypes about warmth and competence*. https://osf.io/hrm2t/

Bar-Haim, Y., Ziv, T., Lamy, D., & Hodes, R. M. (2006). Nature and Nurture in Own-Race Face Processing. *Psychological Science*, *17*(2), 159–163. https://doi.org/10.1111/j.1467-9280.2006.01679.x

Barner, D., Brooks, N., & Bale, A. (2011). Accessing the unsaid: The role of scalar alternatives in children's pragmatic inference. *Cognition*, *118*(1), 84–93. https://doi.org/10.1016/j.cognition.2010.10.010

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*(1), 53–58. https://doi.org/10.1111/j.1467-9280.2005.01664.x

Bell, A. (1984). Language Style as Audience Design. *Language in Society*, *13*(2), 145–204.

Bergsieker, H. B., Leslie, L. M., Constantine, V. S., & Fiske, S. T. (2012). Stereotyping by omission: Eliminate the negative, accentuate the positive. *Journal of Personality and Social Psychology*, *102*(6), 1214–1238. https://doi.org/10.1037/a0027717

Bian, L., Leslie, S.-J., & Cimpian, A. (2017). Gender stereotypes about intellectual ability emerge early and influence children's interests. *Science*, *355*(6323), 389–391. https://doi.org/10.1126/science.aah6524

Bian, L., Leslie, S.-J., & Cimpian, A. (2018). Evidence of bias against girls and women in contexts that emphasize intellectual ability. *American Psychologist*, *73*(9), 1139–1153. https://doi.org/10.1037/amp0000427

Bigler, R. S., & Liben, L. S. (2007). Developmental Intergroup Theory: Explaining and Reducing Children's Social Stereotyping and Prejudice. *Current Directions in Psychological Science*, *16*(3), 162–166.

Boseovski, J. J. (2010). Evidence for “Rose-Colored Glasses”: An Examination of the Positivity Bias in Young Children’s Personality Judgments. *Child Development Perspectives*, *4*(3), 212–218. https://doi.org/10.1111/j.1750-8606.2010.00149.x

Cameron, J. A., Alvarez, J. M., Ruble, D. N., & Fuligni, A. J. (2001). Children’s Lay Theories About Ingroups and Outgroups: Reconceptualizing Research on Prejudice. *Personality and Social Psychology Review*, *5*(2), 118–128. https://doi.org/10.1207/S15327957PSPR0502_3

Castelli, L., De Amicis, L., & Sherman, S. J. (2007). The loyal member effect: On the preference for ingroup members who engage in exclusive relations with the ingroup. *Developmental Psychology*, *43*(6), 1347–1359. https://doi.org/10.1037/0012-1649.43.6.1347

Castelli, L., De Dea, C., & Nesdale, D. (2008). Learning Social Attitudes: Children’s Sensitivity to the Nonverbal Behaviors of Adult Models During Interracial Interactions. *Personality and Social Psychology Bulletin*, *34*(11), 1504–1513. https://doi.org/10.1177/0146167208322769

Cheryan, S., & Bodenhausen, G. V. (2000). When Positive Stereotypes Threaten Intellectual Performance: The Psychological Hazards of “Model Minority” Status. *Psychological Science*, *11*(5), 399–402. https://doi.org/10.1111/1467-9280.00277

Chestnut, E. K., & Markman, E. M. (2016). Are Horses Like Zebras, or Vice Versa? Children’s Sensitivity to the Asymmetries of Directional Comparisons. *Child Development*, *87*(2), 568–582. https://doi.org/10.1111/cdev.12476

Chestnut, E. K., & Markman, E. M. (2018). “Girls Are as Good as Boys at Math” Implies That Boys Are Probably Better: A Study of Expressions of Gender Equality. *Cognitive Science*, *42*(7), 2229–2249. https://doi.org/10.1111/cogs.12637

Chestnut, E. K., Zhang, M. Y., & Markman, E. M. (2021). “Just as good”: Learning gender stereotypes from attempts to counteract them. *Developmental Psychology*, *57*(1), 114–125. https://doi.org/10.1037/dev0001143

Cimpian, A. (2013). Generic Statements, Causal Attributions, and Children’s Naive Theories. In M. R. Banaji & S. A. Gelman (Eds.), *Navigating the Social World* (pp. 269–274). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199890712.003.0049

Cimpian, A., Arce, H.-M. C., Markman, E. M., & Dweck, C. S. (2007). Subtle Linguistic Cues Affect Children’s Motivation. *Psychological Science*, *18*(4), 314–316. https://doi.org/10.1111/j.1467-9280.2007.01896.x

Cimpian, A., & Markman, E. M. (2008). Preschool children’s use of cues to generic meaning. *Cognition*, *107*(1), 19–53. https://doi.org/10.1016/j.cognition.2007.07.008

Cimpian, A., & Markman, E. M. (2011). The Generic/Nongeneric Distinction Influences How Children Interpret New Information About Social Others: Generic/Nongeneric Distinction. *Child Development*, *82*(2), 471–492. https://doi.org/10.1111/j.1467-8624.2010.01525.x

Clark, H. H. (1998). Communal lexicons. In K. Malmkjaer Williams (Ed.), *Context in language learning and language understanding* (pp. 63–68). Cambridge University Press.

Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2007). The BIAS map: Behaviors from intergroup affect and stereotypes. *Journal of Personality and Social Psychology*, *92*(4), 631–648. https://doi.org/10.1037/0022-3514.92.4.631

Cvencek, D., Meltzoff, A. N., & Greenwald, A. G. (2011). Math-Gender Stereotypes in Elementary School Children: Gender Stereotypes. *Child Development*, *82*(3), 766–779. https://doi.org/10.1111/j.1467-8624.2010.01529.x

Cvencek, D., Nasir, N. S., O'Connor, K., Wischnia, S., & Meltzoff, A. N. (2015). The Development of Math-Race Stereotypes: "They Say Chinese People Are the Best at Math." *Journal of Research on Adolescence*, *25*(4), 630–637. https://doi.org/10.1111/jora.12151

Darley, J. M., & Fazio, R. H. (1980). Expectancy confirmation processes arising in the social interaction sequence. *American Psychologist*, *35*(10), 867–881. https://doi.org/10.1037/0003-066X.35.10.867

Dautel, J. B., & Kinzler, K. D. (2018). Once a French Speaker, Always a French Speaker? Bilingual Children's Thinking About the Stability of Language. *Cognitive Science*, *42*, 287–302. https://doi.org/10.1111/cogs.12572

Dovidio, J. F. (2001). On the Nature of Contemporary Prejudice: The Third Wave. *Journal of Social Issues*, *57*(4), 829–849. https://doi.org/10.1111/0022-4537.00244

Dunham, Y., Baron, A. S., & Banaji, M. R. (2006). From American City to Japanese Village: A Cross-Cultural Investigation of Implicit Race Attitudes. *Child Development*, *77*(5), 1268–1281. https://doi.org/10.1111/j.1467-8624.2006.00933.x

Dupree, C. H., & Fiske, S. T. (2019). Self-presentation in interracial settings: The competence downshift by White liberals. *Journal of Personality and Social Psychology*, *117*(3), 579–604. https://doi.org/10.1037/pspi0000166

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Fiske, S. T. (2015). Intergroup biases: A focus on stereotype content. *Current Opinion in Behavioral Sciences*, *3*, 45–50. https://doi.org/10.1016/j.cobeha.2015.01.010

Fiske, S. T. (2018). Stereotype Content: Warmth and Competence Endure. *Current Directions in Psychological Science*, *27*(2), 67–73. https://doi.org/10.1177/0963721417738825

Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, *11*(2), 77–83. https://doi.org/10.1016/j.tics.2006.11.005

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*(6), 878–902. https://doi.org/10.1037//0022-3514.82.6.878

Gelman, S. (2004). Psychological essentialism in children. *Trends in Cognitive Sciences*, *8*(9), 404–409. https://doi.org/10.1016/j.tics.2004.07.001

Gelman, S. A., Goetz, P. J., Sarnecka, B. W., & Flukes, J. (2008). Generic Language in Parent-Child Conversations. *Language Learning and Development*, *4*(1), 1–31. https://doi.org/10.1080/15475440701542625

Gelman, S. A., & Raman, L. (2003). Preschool Children Use Linguistic Form Class and Pragmatic Cues to Interpret Generics. *Child Development*, *74*(1), 308–325. https://doi.org/10.1111/1467-8624.00537

Giles, H., & Ogay, T. (2007). Communication Accommodation Theory. In B. B. Whaley & W. Samter (Eds.), *Explaining communication: Contemporary theories and exemplars* (pp. 293–310). Lawrence Erlbaum Associates Publishers.

Grice, H. P. (1975). Logic and Conversation. In P. Cole & J. L. Morgan (Eds.), *Speech Acts* (pp. 41–58). Academic Press. https://doi.org/10.1163/9789004368811_003

Grodner, D., & Sedivy, J. C. (2011). The Effect of Speaker-Specific Information on Pragmatic Inferences. In E. A. Gibson & N. J. Pearlmutter (Eds.), *The Processing and Acquisition of Reference* (pp. 239–272). The MIT Press. https://doi.org/10.7551/mitpress/9780262015127.003.0010

Hagá, S., & Olson, K. R. (2017). Knowing-it-all but still learning: Perceptions of one's own knowledge and belief revision. *Developmental Psychology*, *53*(12), 2319–2332. https://doi.org/10.1037/dev0000433

Hazell, V., & Clarke, J. (2008). Race and Gender in the Media: A Content Analysis of Advertisements in Two Mainstream Black Magazines. *Journal of Black Studies*, *39*(1), 5–21. https://doi.org/10.1177/0021934706291402

Heyman, G. D., Gee, C. L., & Giles, J. W. (2003). Preschool Children's Reasoning About Ability. *Child Development*, *74*(2), 516–534. https://doi.org/10.1111/1467-8624.7402013

Horowitz, A. C., & Frank, M. C. (2016). Children's Pragmatic Inferences as a Route for Learning About the World. *Child Development*, *87*(3), 807–819. https://doi.org/10.1111/cdev.12527

Huang, Y. T., & Snedeker, J. (2009). Semantic meaning and pragmatic interpretation in 5-year-olds: Evidence from real-time spoken language comprehension. *Developmental Psychology*, *45*(6), 1723–1739. https://doi.org/10.1037/a0016704

Ikeda, A., Kobayashi, T., & Itakura, S. (2018). Sensitivity to linguistic register in 20-month-olds: Understanding the register-listener relationship and its abstract rules. *PLOS ONE*, *13*(4), e0195214. https://doi.org/10.1371/journal.pone.0195214

Jaxon, J., Lei, R. F., Shachnai, R., Chestnut, E. K., & Cimpian, A. (2019). The Acquisition of Gender Stereotypes about Intellectual Ability: Intersections with Race. *Journal of Social Issues*, *75*(4), 1192–1215. https://doi.org/10.1111/josi.12352

Johnson, P. O., & Neyman, J. (1936). Tests of certain linear hypotheses and their application to some educational problems. *Statistical Research Memoirs*, *1*, 57–93.

Kao, G. (2000). Group Images and Possible Selves Among Adolescents: Linking Stereotypes to Expectations by Race and Ethnicity. *Sociological Forum*, *15*(3), 407–430. https://doi.org/10.1023/A:1007572209544

Kelly, D. J., Liu, S., Lee, K., Quinn, P. C., Pascalis, O., Slater, A. M., & Ge, L. (2009). Development of the other-race effect during infancy: Evidence toward universality? *Journal of Experimental Child Psychology*, *104*(1), 105–114. https://doi.org/10.1016/j.jecp.2009.01.006

King, R., & Kinzler, K. (2022). *The Development of Social Group Concepts and Stereotypes about Wealth and Urban-Rural Geography* [Poster]. Cognitive Development Society Bi-Ennial Conference, Madison, WI.

Kinzler, K. D., Shutts, K., DeJesus, J., & Spelke, E. S. (2009). Accent Trumps Race in Guiding Children's Social Preferences. *Social Cognition*, *27*(4), 623–634. https://doi.org/10.1521/soco.2009.27.4.623

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, *44*(4), 978–990. https://doi.org/10.3758/s13428-012-0210-4

Lee, T. L., & Fiske, S. T. (2006). Not an outgroup, not yet an ingroup: Immigrants in the Stereotype Content Model. *International Journal of Intercultural Relations*, *30*(6), 751–768. https://doi.org/10.1016/j.ijintrel.2006.06.005

Lei, R. F., Leshin, R. A., & Rhodes, M. (2020). The Development of Intersectional Social Prototypes. *Psychological Science*, *31*(8), 911–926. https://doi.org/10.1177/0956797620920360

Leslie, S.-J., Cimpian, A., Meyer, M., & Freeland, E. (2015). Expectations of brilliance underlie gender distributions across academic disciplines. *Science*, *347*(6219), 262–265. https://doi.org/10.1126/science.1261375

Lin, M. H., Kwan, V. S. Y., Cheung, A., & Fiske, S. T. (2005). Stereotype Content Model Explains Prejudice for an Envied Outgroup: Scale of Anti-Asian American Stereotypes. *Personality and Social Psychology Bulletin*, *31*(1), 34–47. https://doi.org/10.1177/0146167204271320

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, *47*(4), 1122–1135. https://doi.org/10.3758/s13428-014-0532-5

MacWhinney, B., & Snow, C. (1990). The Child Language Data Exchange System: An update. *Journal of Child Language*, *17*(2), 457–472. https://doi.org/10.1017/S0305000900013866

Mandalaywala, T. M., Ranger-Murdock, G., Amodio, D. M., & Rhodes, M. (2019). The Nature and Consequences of Essentialist Beliefs About Race in Early Childhood. *Child Development*, *90*(4). https://doi.org/10.1111/cdev.13008

Master, A., Meltzoff, A. N., & Cheryan, S. (2021). Gender stereotypes about interests start early and cause gender disparities in computer science and engineering. *Proceedings of the National Academy of Sciences*, *118*(48), e2100030118. https://doi.org/10.1073/pnas.2100030118

Mastro, D. (2017). Race and Ethnicity in US Media Content and Effects. In D. Mastro, *Oxford Research Encyclopedia of Communication*. Oxford University Press. https://doi.org/10.1093/acrefore/9780190228613.013.122

McGlothlin, H., Killen, M., & Edmonds, C. (2005). European-American children's intergroup attitudes about peer relationships. *British Journal of Developmental Psychology*, *23*(2), 227–250. https://doi.org/10.1348/026151005x26101

Mckown, C., Gregory, A., & Weinstein, R. S. (2010). Expectations, stereotypes, and self-fulfilling prophecies in classroom and school life. In *Handbook of research on schools, schooling, and human development* (pp. 256–274).

Meyer, M., Cimpian, A., & Leslie, S.-J. (2015). Women are underrepresented in fields where success is believed to require brilliance. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00235

Moty, K., & Rhodes, M. (2021). The Unintended Consequences of the Things We Say: What Generic Statements Communicate to Children About Unmentioned Categories. *Psychological Science*, *32*(2), 189–203. https://doi.org/10.1177/0956797620953132

Muramatsu, N., & Chin, M. H. (2022). Battling Structural Racism Against Asians in the United States: Call for Public Health to Make the "Invisible" Visible. *Journal of Public Health Management and Practice*, *28*(Supplement 1), S3–S8. https://doi.org/10.1097/PHH.0000000000001411

Museus, S. D., & Kiang, P. N. (2009). Deconstructing the model minority myth and how it contributes to the invisible minority reality in higher education research. *New Directions for Institutional Research*, *2009*(142), 5–15. https://doi.org/10.1002/ir.292

National Center for Education Statistics. (2019). *Status and Trends in the Education of Racial and Ethnic Groups*. https://nces.ed.gov/programs/raceindicators/indicator_rfa.asp

Nesdale, D. (2001). Language and the Development of Children's Ethnic Prejudice. *Journal of Language and Social Psychology*, *20*(1–2), 90–110. https://doi.org/10.1177/0261927X01020001005

Newheiser, A.-K., Dunham, Y., Merrill, A., Hoosain, L., & Olson, K. R. (2014). Preference for high status predicts implicit outgroup bias among children from low-status groups. *Developmental Psychology*, *50*(4), 1081–1090. https://doi.org/10.1037/a0035054

Noveck, I. A. (2001). When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition*, *78*(2), 165–188. https://doi.org/10.1016/S0010-0277(00)00114-1

Olson, K. R., Shutts, K., Kinzler, K. D., & Weisman, K. G. (2012). Children Associate Racial Groups With Wealth: Evidence From South Africa. *Child Development*, *83*(6), 1884–1899. https://doi.org/10.1111/j.1467-8624.2012.01819.x

Papafragou, A., & Musolino, J. (2003). Scalar implicatures: Experiments at the semantics-pragmatics interface. *Cognition*, *86*(3), 253–282. https://doi.org/10.1016/S0010-0277(02)00179-8

Park, J. Z., Martinez, B. C., Cobb, R., Park, J. J., & Wong, E. R. (2015). Exceptional Outgroup Stereotypes and White Racial Inequality Attitudes toward Asian Americans. *Social Psychology Quarterly*, *78*(4), 399–411. https://doi.org/10.1177/0190272515606433

Pauker, K., Ambady, N., & Apfelbaum, E. P. (2010). Race Salience and Essentialist Thinking in Racial Stereotype Development: Racial Stereotype Development. *Child Development*, *81*(6), 1799–1813. https://doi.org/10.1111/j.1467-8624.2010.01511.x

Perszyk, D. R., Lei, R. F., Bodenhausen, G. V., Richeson, J. A., & Waxman, S. R. (2019). Bias at the intersection of race and gender: Evidence from preschool-aged children. *Developmental Science*, *22*(3), e12788. https://doi.org/10.1111/desc.12788

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Raabe, T., & Beelmann, A. (2011). Development of Ethnic, Racial, and National Prejudice in Childhood and Adolescence: A Multinational Meta-Analysis of Age Differences: Age Differences in Prejudice. *Child Development*, *82*(6), 1715–1737. https://doi.org/10.1111/j.1467-8624.2011.01668.x

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

Roussos, G., & Dunham, Y. (2016). The development of stereotype content: The use of warmth and competence in assessing social groups. *Journal of Experimental Child Psychology*, *141*, 133–144. https://doi.org/10.1016/j.jecp.2015.08.009

Ruble, D. N., & Dweck, C. S. (1995). Self-perceptions, person conceptions, and their development. In *Social development.* (pp. 109–139). Sage Publications, Inc.

Rutland, A., Killen, M., & Abrams, D. (2010). A New Social-Cognitive Developmental Perspective on Prejudice: The Interplay Between Morality and Group Identity. *Perspectives on Psychological Science*, *5*(3), 279–291. https://doi.org/10.1177/1745691610369468

Sachs, J., & Devin, J. (1976). Young children's use of age-appropriate speech styles in social interaction and role-playing. *Journal of Child Language*, *3*(1), 81–98. https://doi.org/10.1017/S030500090000132X

Sanchez, A., Meylan, S. C., Braginsky, M., MacDonald, K. E., Yurovsky, D., & Frank, M. C. (2019). childes-db: A flexible and reproducible interface to the child language data exchange system. *Behavior Research Methods*, *51*(4), 1928–1941. https://doi.org/10.3758/s13428-018-1176-7

Saxton, M. (2009). *The Inevitability of Child Directed Speech* (S. Foster-Cohen, Ed.; pp. 62–86). Palgrave Macmillan UK. https://doi.org/10.1057/9780230240780_4

Shatz, M., & Gelman, R. (1973). The Development of Communication Skills: Modifications in the Speech of Young Children as a Function of Listener. *Monographs of the Society for Research in Child Development*, *38*(5), 1. https://doi.org/10.2307/1165783

Shu, Y., Hu, Q., Xu, F., & Bian, L. (2022). Gender stereotypes are racialized: A cross-cultural investigation of gender stereotypes about intellectual talents. *Developmental Psychology*, *58*(7), 1345–1359. https://doi.org/10.1037/dev0001356

Shutts, K. (2015). Young Children's Preferences: Gender, Race, and Social Status. *Child Development Perspectives*, *9*(4), 262–266. https://doi.org/10.1111/cdep.12154

Shutts, K., Kinzler, K. D., Katz, R. C., Tredoux, C., & Spelke, E. S. (2011). Race preferences in children: Insights from South Africa: Social preferences in South Africa. *Developmental Science*, *14*(6), 1283–1291. https://doi.org/10.1111/j.1467-7687.2011.01072.x

Shutts, K., Roben, C. K. P., & Spelke, E. S. (2013). Children's Use of Social Categories in Thinking About People and Social Relationships. *Journal of Cognition and Development*, *14*(1), 35–62. https://doi.org/10.1080/15248372.2011.638686

Sierksma, J., Brey, E., & Shutts, K. (2022). Racial Stereotype Application in 4-to-8-Year-Old White American Children: Emergence and Specificity. *Journal of Cognition and Development*, *23*(5), 660–685. https://doi.org/10.1080/15248372.2022.2090945

Stiller, A. J., Goodman, N. D., & Frank, M. C. (2015). Ad-hoc Implicature in Preschool Children. *Language Learning and Development*, *11*(2), 176–190. https://doi.org/10.1080/15475441.2014.927328

Tajfel, H. (1970). Experiments in Intergroup Discrimination. *Scientific American*, *223*(5), 96–103. JSTOR.

Tajfel, H., & Turner, J. C. (1979). An integrative theory of inter-group conflict. In W. G. Austin & Worchel (Eds.), *The social psychology of inter-group relations* (pp. 33–47). Brooks/Cole.

Tukachinsky, R., Mastro, D., & Yarchi, M. (2015). Documenting Portrayals of Race/Ethnicity on Primetime Television over a 20-Year Span and Their Association with National-Level Racial/Ethnic Attitudes: TV Portrayals and National-Level Attitudes. *Journal of Social Issues*, *71*(1), 17–38. https://doi.org/10.1111/josi.12094

U.S. Census Bureau. (2020). *Inequalities Persist Despite Decline in Poverty for All Major Race and Hispanic Origin Groups*. https://www.census.gov/library/stories/2020/09/poverty-rates-for-blacks-and-hispanics-reached-historic-lows-in-2019.html

U.S. Government Accountability Office. (2019). *K-12 Education: Public High Schools with More Students in Poverty and Smaller Schools Provide Fewer Academic Offerings to Prepare for College*. U.S. Government Accountability Office.

Wagner, L., Greene-Havas, M., & Gillespie, R. (2010). Development in children's comprehension of linguistic register. *Child Development*, *81*(6), 1678–1686. https://doi.org/10.1111/j.1467-8624.2010.01502.x

Weisbuch, M., Pauker, K., & Ambady, N. (2009). The Subtle Transmission of Race Bias via Televised Nonverbal Behavior. *Science*, *326*(5960), 1711–1714. https://doi.org/10.1126/science.1178358

Yussen, S. R., & Kane, P. (1985). Children's concept of intelligence. In S. R. Yussen (Ed.), *The growth of reflection in children*. Academic Press.