Developmental Science

WILEY

Developmental Science

RESEARCH ARTICLE

The Emergence of Belief Attribution and Dehumanization Are Associated

Wen Zhou[1] | Brian Hare[2,3,4]

[1]Division of Social Sciences, Duke Kunshan University, Jiangsu, China | [2]Department of Evolutionary Anthropology, Duke University, Durham, North Carolina, USA | [3]Center for Cognitive Neuroscience, Duke University, Durham, North Carolina, USA | [4]Department of Psychology and Neuroscience, Duke University, Durham, North Carolina, USA

Correspondence: Wen Zhou (wen.zhou@duke.edu)

Received: 7 August 2025 | **Revised:** 15 February 2026 | **Accepted:** 19 February 2026

Keywords: dehumanization | intergroup bias | moral development | theory-of-mind

ABSTRACT

Dehumanization is hypothesized to involve denying others a fully human mind. We tested its proposed link with theory-of-mind development in 3- to 6-year-olds (total $N = 247$) using a minimal group paradigm framed as a competition. Across two experiments, only children who understood false beliefs rated the outgroup as less human than the ingroup, although they liked the outgroup less regardless of their theory-of-mind performance. As theory-of-mind development advanced, outgroup dehumanization increased, with intent to harm the outgroup only being associated with dehumanization among children who also understood second-order beliefs ($n = 51$). However, the strength of this relationship remains uncertain since the effect became marginally significance after controlling for intergroup liking. These results provide initial support for theory-of-mind abilities being related to the development of dehumanization. They also point to the potential for intervention during early childhood before an association between dehumanization and aggression forms.

Introduction

Recognizing members of one's own species is a fundamental social skill. At birth, human infants recognize human faces and distinguish them from nonhumans (Heron-Delaney et al. 2011). Infants are attracted to human eyes and seek out eye contact as they quickly begin contingent social smiling and vocalizing (Farroni et al. 2002; Goldstein et al. 2009; Grossmann 2017). But by an infant's first birthday, they have begun a revolution in their understanding of others (Tomasello 2019). Infants not only recognize the physical features of humans, but also their internal mental states. By 9–12 months of age, they begin recognizing in other humans what they can perceive, intend and know (Woodward et al. 2001). They also readily overascribe mental states to a wide range of nonliving things in response to minimal cues of animacy, but soon more selectively specify their attributions to humans (Goldman and Poulin-Dubois 2024; Meltzoff 1995). This development culminates in early childhood with "full blown" representational theory-of-mind, the understanding that others hold knowledge and beliefs that can differ from one's own or reality (Gopnik and Astington 1988; Wellman et al. 2001). But it remains a question if the development of mind understanding influences children's recognition of humanness in different individuals or groups.

Theory of mind in humans, defined as the understanding and attribution of mental states to oneself and others, reaches levels of complexity not observed in other species (Tomasello and Rakoczy 2003). While some nonhumans, like other great apes, are capable of first-order attributions (e.g. reasoning about another individual's belief; Hare et al. 2000; Hare et al. 2001; Krupenye et al. 2016; MacLean and Hare 2012), only humans have the

Summary

- The emergence of dehumanization is associated with the development of theory-of-mind. Only children with belief attribution capacities consider the outgroup as less human.
- Children with more advanced belief attribution demonstrate stronger dehumanization, reflected in lower humanness ratings for the outgroup.
- In preschool-aged children (3- to 6-year-olds), dehumanization of the outgroup is not associated with verbal ability or age.
- In preschool-aged children, dehumanization and intergroup preference appear distinguishable by the relationship of dehumanization to the development of belief attribution abilities.

ability to make more complicated recursive inferences (Dodell-Feder et al. 2013; Hare 2011). This ability extends beyond the individual to groups. Only humans think about the thoughts of groups and how different groups view each other (e.g., Kteily et al. 2016). Our theory-of-mind which makes us uniquely human also allows us to see human minds in others (Hare 2011, 2017). We humanize each other, nonhuman animals, natural forces and even inanimate objects when we see them as mental agents (Epley et al. 2007; Zhou et al. 2024). This flexible capacity to attribute minds is proposed to facilitate dehumanization as well, by enabling humans to selectively attribute or withhold forms of social cognition to the minds of others (Gray et al. 2007; Harris 2017; Zhou 2022).

Dehumanization is the psychological process of denying humanness or a degree of humanness to certain individuals or groups (Kteily and Landry 2022). In adults, dehumanization has been conceived to expresses itself in a variety of forms, and often has been theorized to involve perceiving the mental life of the dehumanized as reduced or less complex (Haslam et al. 2006; Waytz and Schroeder 2014; Kteily and Landry 2022). When other humans are not fully attributed a human mind, or at least denied aspects of it, they are frequently likened to objects or nonhumans that do not possess human-unique psychology (Gray et al. 2012; Harris and Fiske 2011; Loughnan et al. 2009; Morera et al. 2018). This mind denial occurs readily in intergroup contexts. For instance, when adults are shown a face continuum ranging from human to inanimate doll, they are more likely to perceive the human faces as having a mind, and this mind perception occurs more readily when the human faces are perceived as ingroup rather than outgroup (Hackel et al. 2014). As a core aspect of dehumanization, mind denial plays a central role in moral disengagement which may drive the link between dehumanization and harm (Gray et al. 2012; Hare 2017; Harris 2017). Seeing humanity endorses the recognition of suffering and moral considerations, while dehumanization can justify indifference, aggression or cruelty (Čehajić et al. 2009; Haslam et al. 2012).

Although the developmental trajectory of dehumanization remains underexplored, existing evidence suggests a potential link with the development of mind attribution. Theory-of-mind develops markedly by the end of preschool years (Wellman et al. 2001). Notably, around this same age, children begin to exhibit adult-like patterns of intergroup bias in mind perception, attributing richer mental states to ingroup than to outgroup members (McLoughlin and Over 2017). As theory-of-mind comes online, children are also more likely to favor their ingroup in altruistic sharing and are more inclined to punish the outgroup (Yu et al. 2024). Children as young as five years old, when given a graphic scale, are also more likely to rank ingroup members as more human than the outgroup members (Mcloughlin et al. 2018; Zhou and Hare 2022). The same children are also more likely to indicate an intent to harm outgroup members when they dehumanize them (Zhou and Hare 2022). The observed link between outgroup dehumanization and harmful intent appears robust across school-age children and adults, and it persists even when controlling for affective prejudice (Bruneau et al. 2020; Zhou and Hare 2022). Indeed, while ingroup favoritism motivates selectively benefiting the ingroup, it does not always predict a willingness to harm outgroups (Brewer 2001, 2007; Cameron et al. 2003).

Despite the potential developmental link between theory-of-mind abilities and out group dehumanization, there has never been a direct test designed to examine if their emergence is related. Individual variability in the development of humans' unique mentalizing abilities provide the potential for a powerful test. Differences across individuals have been linked to a variety of problems solving and prosocial preferences that allow for unique forms of cooperation and cultural learning (Imuta et al. 2016). Studies of emergence and maturational patterns have led to the overall picture that mind understanding is fundamental to human norms, sharing, helping and collaboration (Tomasello 2018). What is needed now is a test of whether individual differences in theory-of-mind are also linked to dehumanization in young children.

Here, we test the predicted association by examining whether children's capacity for mind attribution is related to a tendency to dehumanize a novel outgroup. We used a minimal group manipulation framed as intergroup competition, since competitive interactions reliably induces dehumanizing tendencies in adults and 5- to 12-year-old children (Fiske 2013; Zhou and Hare 2022). In Experiment 1, we test young children from 3- to 6-year-olds since theory-of-mind abilities, and belief attribution in particular, are highly variable at these ages (Liu et al. 2008). We measured dehumanization using graphic scales that children as young as five years of age intuitively understand after a brief introduction (Zhou and Hare 2022). While these perceptual scales do not test between the attribution of different complex social-emotional states (e.g., uniquely human characteristics such as pride, shame, etc., Haslam 2006), they likely measure a more basic or incipient understanding of humanness that likely emerges earlier in development and does not require an understanding of subtle differences in complex emotional states or advanced vocabulary (Hui and Kung 2025; Zhou and Hare 2022). However, we did include a measure of verbal ability to provide an additional control for task comprehension since we are testing a relatively wide range of ages (Osterhaus and Bosacki 2022). In Experiment 2, we replicated Experiment 1 with naïve sample of 3- to 4-year-olds. We focused on a narrower age range to better capture dehumanizing tendencies in children who do and do

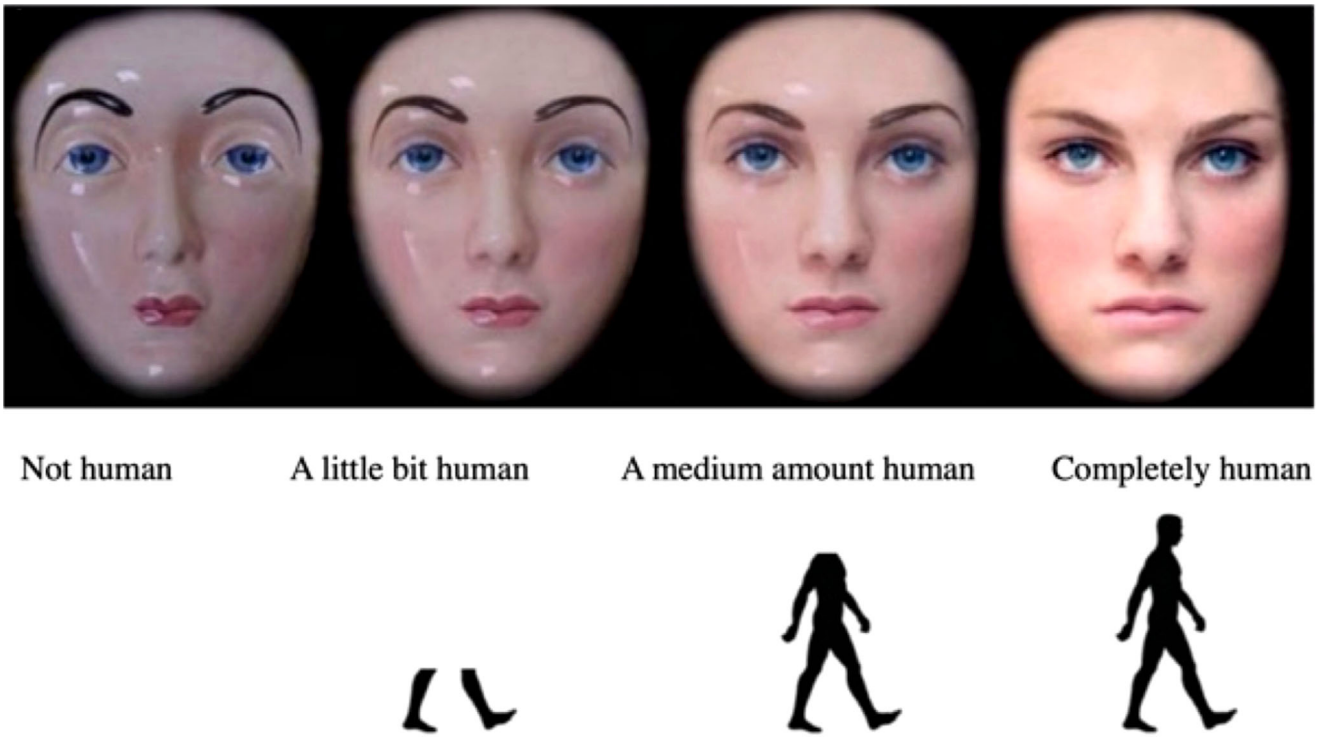

FIGURE 1 | Children were given a 4-point composite scale to rate the humanness of their ingroup and the hypothetical peer outgroup by answering the question: "To what degree do you think kids in the Green/Yellow group fit the definition of being human? Not at all, a little bit, a medium amount, or a lot?".

not show skill attributing beliefs to others. In both experiments, we additionally tested intergroup liking and intentions to harm the outgroup. In Experiment 2, we further included a measure designed to control the perception of ingroup dominance which is associated with dehumanization in adults (Kteily et al. 2015).

If there is a developmental link between mentalizing and dehumanization, then children who succeed on mind attribution tasks will be more likely to rate an outgroup as less human than the ingroup, and children with more advanced mind attribution skills will also show even stronger dehumanizing tendencies. Given the central role of mentalizing in moral engagement, the same children will also be more likely to exhibit harmful intents. Alternatively, mentalizing abilities are either unpaired with or more generally related to other forms of outgroup derogation (Enock et al. 2021). This predicts mentalizing abilities will either be unrelated to outgroup measures or will be similarly associated to our secondary measure of intergroup liking as well as dehumanization. And the role of dehumanization in predicting harm may not be distinct from that of intergroup liking.

1 | Experiment 1

We first conducted an initial cross-sectional test of the proposed link between mind attribution capacities and dehumanization in preschoolers from a local kindergarten. Using a minimal group paradigm, we compared participants' ratings of humanness of their ingroup and the outgroup. Since preschoolers are mostly non-literate, we deployed graphic scales (Figure 1) which have raised the possibility that children are already capable of a form of blatant denial of humanness (Mcloughlin et al. 2018; Zhou and Hare 2022). Graphic scales provide a promising tool since 93.17% of child participants in Zhou and Hare (2022), without training, were able to spontaneously match items on these scales with corresponding levels of humanness and showed rates of dehumanization resembling adults tested with the same scales (Zhou and Hare 2022).

To explore if theory-of-mind development plays a role in facilitating dehumanization, we also assessed participants' intent to cause outgroup suffering. Participants were given a chance to decide how spicy the outgroup's food would be. In addition, while a range of evidence suggests blatant dehumanization is distinct from outgroup dislike (Bruneau et al. 2020; Mcloughlin et al. 2018; Waytz and Epley 2012), dehumanization is not affect-neutral and likely interacts heavily with dislike (Fincher et al. 2018). To help determine if any association between theory-of-mind development and our measures of dehumanization are signs of a more general relationship between intergroup preferences and mentalizing (Enock et al. 2021), we also included measures of liking.

2 | Methods

2.1 | Participants

We tested 117 children who were 3 to 6-years of age (Mean age = 57.72 months, *SD* = 11.94, range = 38–81 months; 61 boys, 56 girls; All Asians and native Mandarin Chinese speakers). All participants were recruited from a kindergarten situated in

Kunshan, China (age and gender information was obtained from the kindergarten). Since the relationship between dehumanization and theory-of-mind development has never been tested in young children, an a priori estimation of effect size is unfeasible. Instead of power analysis, we conducted sensitivity analyses and reported the results in the Supplemental Information (Tables S1 and S2).

An additional nine participants were tested but excluded from all analysis because they did not complete the testing ($n = 6$) or failed the introductory session designed to assess children's understanding of humanness ($n = 3$). Otherwise, because we administered a battery of tasks, all participants received the tests in the same order (e.g. Herrmann et al. 2007; Wobber et al. 2014). Participants were introduced to the group exercise and the competitive context, tested on the three theory of mind tasks, introduced to the use of our scales and then presented with the humanness, intergroup liking, and harm scales. Finally, they were tested for verbal ability, which is the most common control measure of general cognitive ability typically paired with theory of mind tasks (Milligan et al. 2007; Osterhaus and Bosacki 2022).

2.2 | Procedure

Participants were tested individually. All experiments were conducted in a designated room of the kindergarten familiar to participants. Two experimenters who were not aware of the research objectives and hypotheses administered the tasks with each child in Mandarin Chinese. The main experimenter (E1) led the participant through the tasks, and repeated participants' responses loudly. The second experimenter (E2) wrote down the participant's responses and audio-recorded each session. After the experiments, E2 double-checked that E1's vocal report of the participants' responses matched the recording coding on each data sheet. On the handful of trials where the written and audio record disagreed in this process, the final scoring was made based on the audio record.

Group Membership: After being seated, E1 informed the child they were participating in a competition between children of their same age. Adapting the minimal group paradigm designed by Tajfel et al. (1971), E1 then asked each participant to choose between a green and a yellow circle and informed them that they would join the team corresponding to their chosen color. Afterward, E1 placed a wristband of the same color on the participant and explained that their team would compete against another team with each team member trying to answer more questions than members of the other team. E1 then confirmed the child remembered which team they had joined using two memory check questions.

Theory-of-Mind: To create the perception of a real competition, participants were told the competition was about listening to and remembering stories. In actuality, the stories were designed to test their theory-of-mind capacities. While theory-of-mind involves multiple components, we focused on perspective taking as it represents children's ability to decouple their own perspective from that of another and attributes autonomous minds to others. To assess this skill, we assessed children's understanding of false beliefs and hidden emotions (e.g. interpreting others desires and emotions commonly requires taking into account their beliefs; Milligan et al. 2007). Although children are often characterized as having a "full-blown" theory-of-mind once they reason about others' beliefs and propositional attitudes, the ability continues to develop as children begin to succeed with second-order inferences and beyond (Butterfill and Apperly 2013; Miller 2009). Thus, to capture meaningful individual differences, we used both first- and second-order false belief tasks. The measure of emotion detection is incorporated in the second-order task.

In the first-order task, E1 conducted the standard Sally-Anne test (Baron-Cohen et al. 1985) using five illustrations depicting the two characters and their actions. The two characters were given the Chinese names, "Fangfang" and "Lili". Participants were told Fangfang put a ball into a basket with an opaque cap, and then she left the room. Then Lili moved the ball secretly out of the basket, put it into a box and closed the box. After recounting the ball's movement in the story, E1 assessed the participant's comprehension of the ball's position. Since this was the first story presented, it is possible that participants of this age were not yet fully attentive. Thus, if the participant answered incorrectly, the experimenter repeated the story and checked for comprehension again. All participants passed the comprehension checks. To evaluate first-order false belief attribution, E1 then showed an illustration depicting the return of Fangfang, and asked, "where will Fangfang look to find the ball?". This critical belief question, like all subsequent test questions, was not repeated. Participants were scored as correct if they responded with the location Fangfang had originally hidden the ball even though they knew Lili had moved the ball while Fangfang was out of the room.

The second-order "surprise" task was adapted from the birthday puppy story of Sullivan et al. (1994). Using seven illustrations that depict the characters and scenarios, E1 recounted the story of a mother trying to surprise her son for his birthday. The participants were told a mother intends to surprise her son by misleading him about the type of birthday gift he will receive. The mother misinforms her son by telling him he will be receiving a gift he did not want. The son shows a happy reaction, indicated with a smiley face. To test if the participant could distinguish between the son's expressed and actual feelings, they were asked both, "What emotion does the son express?" and "How does the son actually feel?" Children needed to answer both correctly to be coded as "correct". E1 then finished the story and participants learned the son accidentally sees the real birthday gift without the mother being present to observe the discovery. As in the first-order task, the experimenter conducted comprehension checks by asking standardized questions.

E1 then evaluated the participants' first-order false belief attribution by asking, "Does the mom know her son sees the gift?". Participants were scored as correct if they answered "No". E1 then evaluated the participants' second-order false belief attribution by relating to each child that the son's grandmother calls the mother to ask if he knows what he is getting for his birthday. E1 then asked each participant, "What would the mom say?" and then asked them to explain "Why would the mom say that?". Participants were considered as demonstrating second-order belief attribution if they indicated 1) the mom did not know the son knew what the real gift was and 2) she was ignorant because she did not see him find the real gift.

Responses to the theory-of-mind questions were coded as "0" (*incorrect*) or "1" (*correct*). This coding scheme allows for the creation of a two-level belief attribution scale: the appearance of first-order belief attribution ("0" = *failed both first-order questions*, "1" = *passed at least one first-order question*), and the appearance of second-order belief attribution ("0" = *failed*, "1" = *passed*). Responses to the emotion detection question were coded as "0" (*incorrect*) or "1" (*correct*).

Humanness: After finishing the theory-of-mind tasks, participants were introduced to the humanness scale (Figure 1) and were asked how human the members of their own team (the ingroup) and the opponent team (the outgroup) were. They were asked to indicate the level of humanness by circling the appropriate level using a 4-point scale that combined the Face Continuum scale and the Human Silhouette scale used previously (Mcloughlin et al. 2018; Zhou and Hare 2022; Figure 1). To reduce the possibility that children's ratings reflected strategic responses aimed at favoring their ingroup, participants were informed that they had already won the game before the rating task began. To minimize social desirability bias, once they made their choice on the paper displaying the scales, they folded and placed it in a box to make clear their choice was anonymous. The responses were coded from 0 = "*not at all*" to 3 = "*a lot*". Based on a similar metaphor-based dehumanization scale used by Kteily et al. (2015), we further obtained the dehumanization score by subtracting the outgroup humanness rating from the ingroup rating. Participants always rated the ingroup first, and then given the scale again to rate the outgroup. In the introductory phase (described below), we ensured that participants understood that the same response options could be applied to evaluate multiple targets.

To ensure comprehension of the humanness scale (Figure 1), we adopted an introductory phase prior to collecting participants humanness ratings which was designed for young children (Mcloughlin et al. 2018). Each participant was introduced to a bar chart (Figure 2A) to learn the four ascending degrees on the scale, "*Not at all*", "*A little bit*", "*A medium amount*", and "*A lot*". Next, participants were exposed to four jugs (Figure 2B) and quickly learned to judge how much liquid was in each jug using the four-point scale. The participants were then given a list of five jugs, two of which were identically filled with a little bit liquid (Figure 2C). The participants were asked to point out the jugs with a little bit of liquid, so they understood the same response option can be used to evaluate multiple entities. Finally, participants were presented with the Silhouette and the Face Continuum, using the four-point scale to judge how much each item on the scales fits the definition of humanness. Children who initially made an error were corrected and asked the questions again (e.g., if a child failed to indicate the first item was not at all human, E1 would inform them). Those who made two mistakes or repeated mistakes were excluded from analysis ($n = 3$).

Intergroup Liking: Immediately after the participant rated the humanness of members of the ingroup or the outgroup, they rated their liking towards the same group. Specifically, the participant responded to the questions "to what degree do you like your green/yellow team?" and "to what degree do you like your opponents, the yellow/green team?" with the 4-point scale illustrated in Figure 2A. This scale took the form of a bar chart with a *not at all* option (no bar, coded as "0") followed by black bars that increased in height to represent *a little* ("1"), *a medium amount* ("2"), and *a lot* ("3"). We subtracted the outgroup rating from the ingroup rating to measure intergroup liking.

Intent to Harm: Participants were asked to determine whether the opposing team would receive spicy or non-spicy hamburgers for dinner. Before choosing, E1 explained that when members of the opposing team ate spicy burgers it "hurt their tummies and made them cry". Participants were then allowed to indicate the spiciness level they chose for the opposition using the 4-point scale illustrated in Figure 2A (from 0 = "*not spicy at all*" to 3 = "*very spicy*").

Verbal Ability (Cronbach's $\alpha = 0.72$): We adapted a battery assessing perceptual reasoning and verbal comprehension that has demonstrated reliability and validity among Chinese children aged 3 to 6 years (Zhan 2012). Using 20 pass/fail items (total score: 0–20), we evaluated object recognition, spatial understanding, categorical and inferential reasoning, and verbal expression. For example, using illustrated scenes, children were asked to locate and name objects, identify spatial relationships, and interpret visual narratives. The test placed particular emphasis on children's understanding of identity within social contexts. In one task, children heard a brief story ("Uncle Zhang visited me. I offered him a cup of tea, and he called me a good kid.") and were then asked questions such as "Who was the guest?", "Who was the good kid?", and "How did I treat the guest?"

2.3 | Analysis

R4.3.0 was used to analyze the data. Humanness and liking ratings, intent to harm, verbal ability and age were treated as continuous. Capacities for first-order and second-order belief attribution, hidden emotion detection and gender were categorical variables. Theory-of-mind measures were considered categorical because each child was given one question in each unique task to demonstrate belief attribution. We included age by months and gender as demographic covariates in all analyses. Because aspects of verbal ability may improve with age (Fuhs and Day 2011), we regressed verbal ability scores on age in months and used the resulting residuals as a covariate in all analyses to avoid colinearity.

To test if dehumanization requires mentalizing capacity, we conducted an ANCOVA to examine the interaction between the capacity for first-order belief attribution (passing/not passing at least one belief attribution question) and the evaluated group's identity (ingroup/outgroup) on humanness ratings. With post hoc comparisons, we evaluated 1) if participants capable of belief attribution were more likely to rate the outgroup as less human than the ingroup, 2) if participants unable to correctly attribute beliefs were capable of dehumanizing and 3) if children with and without belief attribution capacity differed in ingroup or outgroup humanness ratings. We used the same approach again, examining only children capable of first-order belief attribution, to test for a relationship between their abilities for second-order belief attribution and their intergroup humanness ratings.

As a control, using ANCOVA, we compared ratings of ingroup and outgroup liking in participants who did or did not succeed

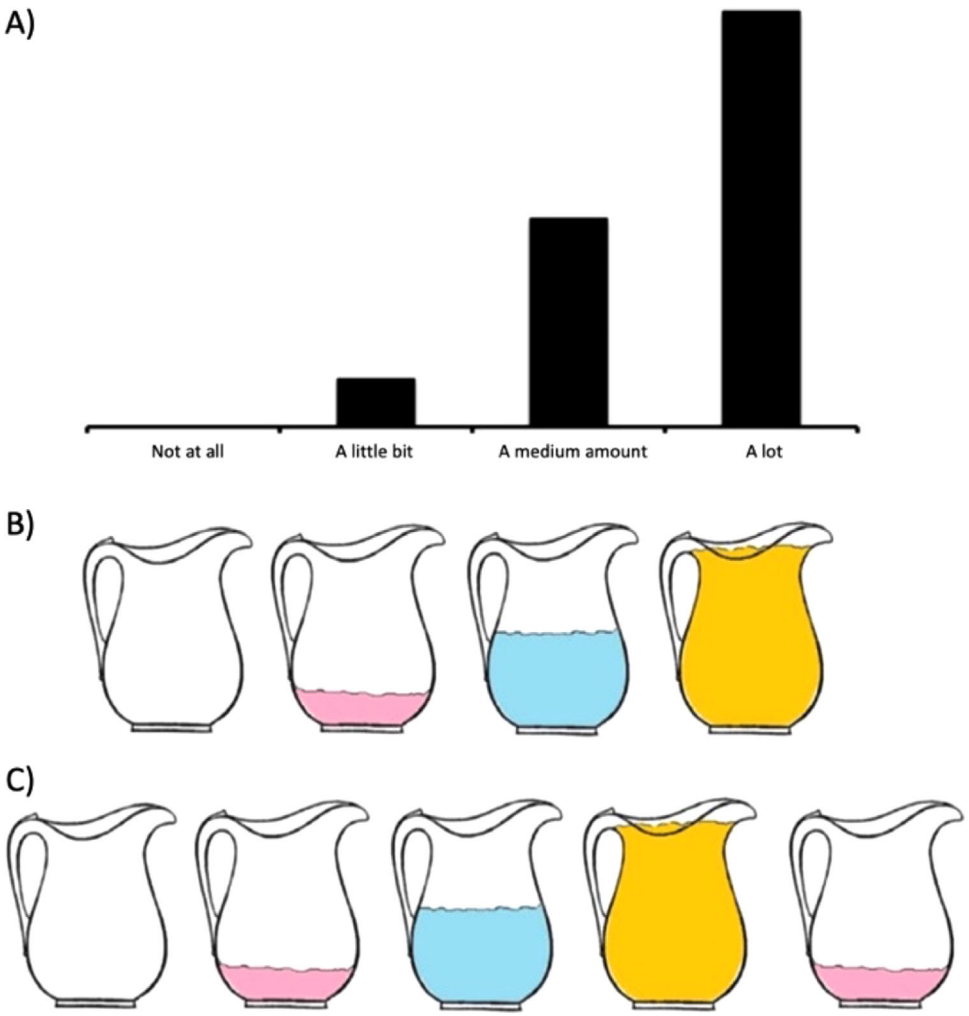

FIGURE 2 | The materials used in the introductory phase to make sure children understand the four-point humanness scale based on Mcloughlin et al. (2018) and Zhou and Hare (2022).

with belief attribution. We also employed linear regression models to estimate the association between intergroup liking and dehumanization, analyzing whether these relationships differed based on the presence of belief attribution capacities.

Finally, to investigate the predictors of harm, we performed linear regression models to test if dehumanization predicted intent to harm and whether this predictive effect varies across children with different levels of belief attribution capacities (no capacity, only capable of first-order reasoning, or capable of second-order reasoning). We further examined the predictive effect controlling for intergroup liking. All tests were two-tailed. Significance values were adjusted using Holm-Bonferroni method.

3 | Results

Among the participants, 94 (47 females, 47 males; Mean age = 60.00 months, *SD* = 11.58 months, range = 38–81 months) passed at least one belief attribution question, demonstrating some theory-of-mind capacities. The other 23 participants (9 females, 14 males; Mean age = 48.39 months, *SD* = 8.48 months, range = 38–75 months) did not show any evidence of belief attribution. Figure 3A illustrates participants' belief attribution capacities by age.

3.1 | The Relationship Between Dehumanization and Theory-of-mind

On humanness ratings, ANCOVA indicated a significant interaction between first-order belief attribution and the evaluated group's identity ($F(1,112) = 8.50$, adjusted $p = 0.02$, $\eta^2 = 0.07$, 95% *CI* [0.01, 1.00]; See Figure 4). Post hoc comparisons showed that participants who demonstrated theory-of-mind capacities gave significantly lower ratings for outgroup humanness (Mean = 1.29, 95% *CI* [1.04, 1.54]) than ingroup humanness (Mean = 2.54, 95% *CI* [2.35, 2.74]; $t(112) = -6.85$, adjusted $p < 0.001$, Cohen's $d = -1.15$, 95% *CI* [−1.45, −0.84]), consistent with the appearance of outgroup dehumanization. Among participants without belief attribution capacities, ratings of outgroup (Mean = 2.17, 95% *CI* [1.69, 2.66]) and ingroup humanness (Mean = 2.35, 95% *CI* [1.90, 2.79]; $t(112) = 1.48$, adjusted $p = 0.14$, Cohen's $d = -0.16$, 95% *CI* [−0.74. 0.42]) did not differ. Children capable

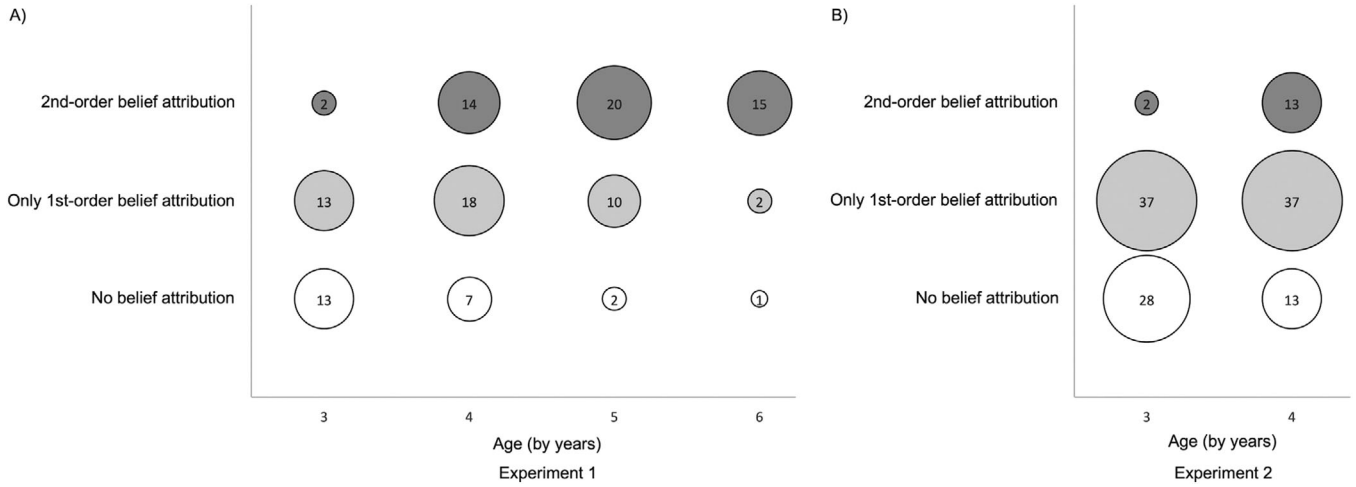

FIGURE 3 | The distribution of participants' belief attribution capacities by age (by years) in Experiments 1 and 2. The bubble size and the number in the center demonstrates the number of participants who correctly answered the second-order attribution question (dark grey bubbles), correctly answered at least one first-order attribution question but failed the second-order question (light grey bubbles), or failed all attribution questions (white bubbles).

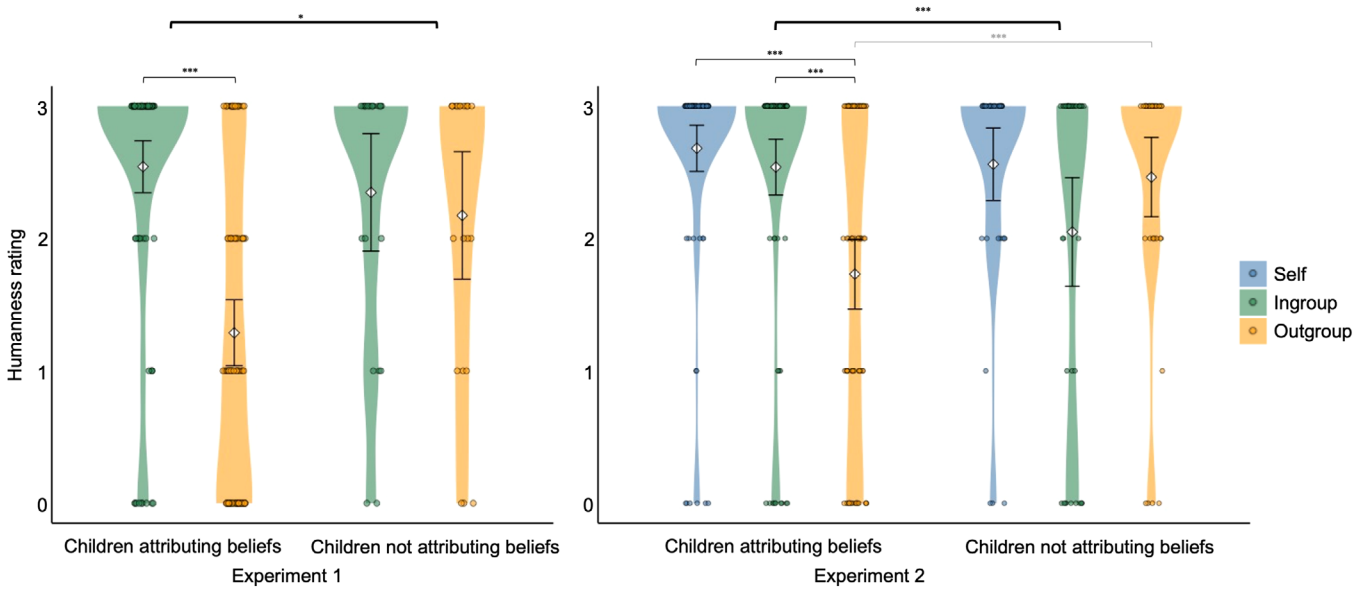

FIGURE 4 | Humanness ratings of the ingroup and the outgroup (Experiments 1 & 2), and participants themselves (Experiment 2) among children with and without first-order belief attribution capacities. Each plot includes the density of ratings (violins), individual data points (dots), means (diamonds) and 95% *CI*s (error bars). *$p < 0.05$, ***$p < 0.001$.

of belief attribution rated the outgroup marginally less human ($t(222) = -1.92$, adjusted $p = 0.056$, Cohen's $d = -0.74$, 95% *CI* [−1.21, −0.27]) in comparison to children who showed no belief attribution, while their ratings of ingroup humanness did not differ ($t(222) = 0.26$, adjusted $p = 0.80$, Cohen's $d = 0.20$, 95% *CI* [−0.26, 0.66]). We did not find any significant interactional effect on humanness ratings between group identity and verbal ability ($F(1,112) = 3.26$, adjusted $p = 0.16$, $\eta^2 = 0.03$, 95% *CI* [0.00, 1.00]), age ($F(1,112) = 3.78$, adjusted $p = 0.16$, $\eta^2 = 0.03$, 95% *CI* [0.00, 1.00]; See the distribution of humanness ratings by age in Figure S1) or gender ($F(1,113) = 0.98$, adjusted $p = 0.32$, $\eta^2 < 0.01$, 95% *CI* [0.00, 1.00]).

On the role of second-order belief attribution, ANCOVA indicated its interaction with the evaluated group's identity on humanness ratings ($F(1,89) = 9.34$, adjusted $p = 0.01$, $\eta^2 = 0.09$, 95%CI [0.02, 1.00]; See Figure 5). Post hoc analyses suggested that the pattern of rating the outgroup (Mean = 0.94, 95% *CI* [0.66, 1.22]) as less human than the ingroup (Mean = 2.65, 95% *CI* [2.42, 2.88]) remains when considering participants who also passed second-order belief attribution ($n = 51$; $t(89) = -8.81$, adjusted $p < 0.001$, Cohen's $d = -1.86$, 95%*CI* [2.32, −1.39]). The effect size in kids with second-order belief understanding is larger than in children who only passed first-order belief attribution ($n = 43$; Outgroup rating: Mean = 1.70, 95% *CI* [1.29, 2.10]; Ingroup rating: Mean

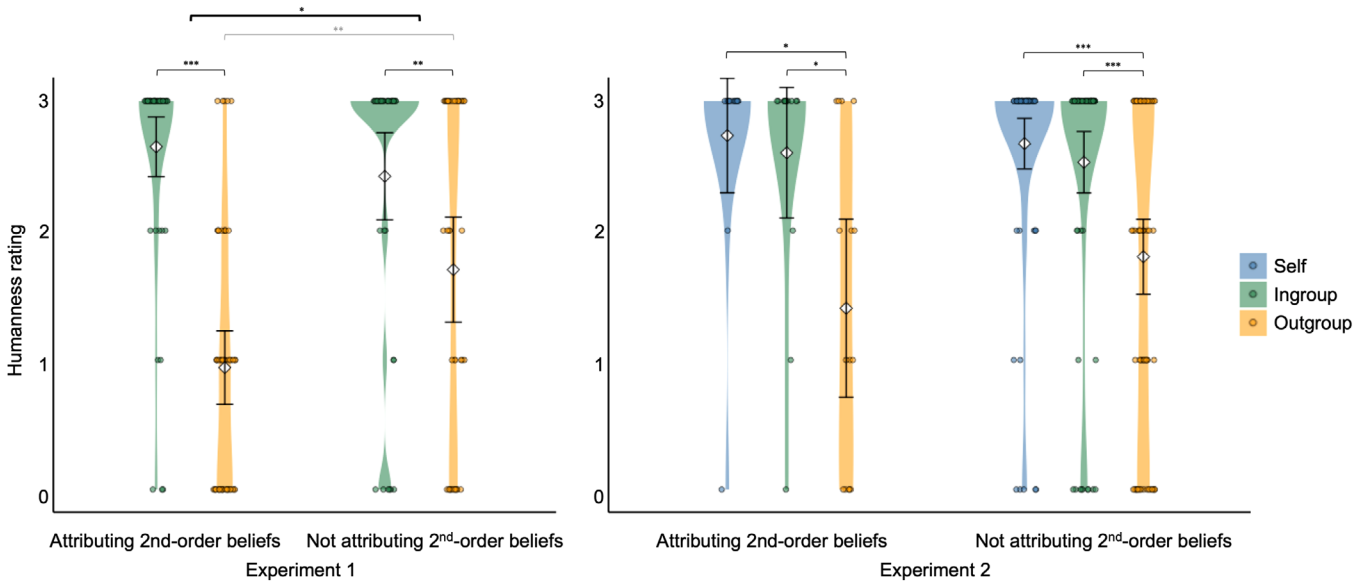

FIGURE 5 | Humanness ratings of the ingroup and the outgroup (Experiments 1 & 2), and participants themselves (Experiment 2) among children who could and could not attribute second-order belief successfully. Each plot includes the density of ratings (violins), individual data points (dots), means (diamonds) and 95% *CI*s (error bars). *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

= 2.42, 95% *CI* [2.08, 2.76]; $t(89) = -2.82$, adjusted $p = 0.006$, Cohen's $d = -0.59$, 95%*CI* [−1.02, −0.16]). Children capable of second-order belief attribution rated the outgroup less human ($t(177) = -2.77$, adjusted $p = 0.006$, Cohen's $d = -0.65$, 95%*CI* [−1.07, −0.23]), compared to children who did not show second-order belief attribution capacity. Children who did or did not attribute second-order belief did not differ in their ratings of ingroup humanness ($t(177) = 0.47$, adjusted $p = 0.64$, Cohen's $d = 0.24$, 95%*CI* [−0.17, 0.65]). We did not find any significant interactional effect on humanness ratings between group identity and verbal ability ($F(1,89) = 1.10$, adjusted $p = 0.89$, $\eta^2 = 0.01$, 95% *CI* [0.00, 1.00]), age ($F(1,89) = 0.003$, adjusted $p = 0.99$, $\eta^2 < 0.01$, 95% *CI* [0.00, 1.00]) or gender ($F(1,89) = 0.36$, adjusted $p = 0.99$, $\eta^2 < 0.01$, 95% *CI* [0.00, 1.00]).

Finally, while results above indicate strong links between belief attribution and outgroup dehumanization, we found no link between the correct evaluation of the supplemental hidden emotions measure and intergroup differences on humanness ratings ($F(1,112) = 1.18$, adjusted $p = 0.56$, $\eta^2 = 0.01$, 95% *CI* [0.00, 1.00]).

3.2 | Distinguishing Between Dehumanization and Intergroup Liking

On liking ratings, ANCOVA indicated a strong effect of the evaluated group's identity ($F(1,111) = 97.09$, adjusted $p < 0.001$, $\eta^2 = 0.47$, 95% *CI* [0.36, 1.00]). We did not find an interaction effect between the evaluated group's identity and belief attribution on ratings of liking ($F(1, 111) = 3.44$, adjusted $p = 0.26$, $\eta^2 = 0.03$, 95% *CI* [0.00, 1.00]; See Figure 6). Post hoc comparisons indicated that liking ratings are significantly lower for outgroup than ingroup in both participants who correctly answered at least one first order belief attribution question (Outgroup: Mean = 1.35, *95% CI* [1.11, 1.60]; Ingroup: Mean = 2.64, *95% CI* [2.50, 2.78]; $t(111) = -8.91$, adjusted $p < 0.001$, Cohen's $d = -1.30$, 95% *CI* [−1.62, −0.99]) and those who did not (Outgroup: Mean = 1.96, 95% *CI* [1.37, 2.55]; Ingroup: Mean = 2.70, *95% CI* [2.39, 3.00]; $t(111) = -3.23$, adjusted $p = 0.002$, Cohen's $d = -0.68$. 95% *CI* [−1.27, −0.08]). We did not see any influence of verbal ability ($F(1, 111) = 0.95$, adjusted $p = 0.36$, $\eta^2 < 0.01$, 95% *CI* [0.00, 1.00]), age ($F(1, 111) = 3.33$, adjusted $p = 0.26$, $\eta^2 = 0.03$, 95% *CI* [0.00, 1.00]) or gender ($F(1, 112) = 1.82$, adjusted $p = 0.36$, $\eta^2 = 0.02$, 95% *CI* [0.00, 1.00]) on intergroup liking.

While humanness and liking ratings demonstrate different patterns with the development of belief attribution, intergroup liking is associated with our dehumanization measure among children with first-order belief attribution ($t(88) = 3.77$, adjusted $p = 0.001$, $\eta^2 = 0.17$, 95% *CI* [0.07, 1.00]) after controlling for verbal ability ($t(88) = 0.99$, adjusted $p = 0.99$, $\eta^2 = 0.01$, 95% *CI* [0.00, 1.00]), age ($t(88) = 1.06$, adjusted $p = 0.99$, $\eta^2 = 0.01$, 95% *CI* [0.00, 1.00]) and gender ($t(88) = 0.01$, $p = 0.99$, $\eta^2 < 0.01$, 95% *CI* [0.00, 1.00]). We did not find this link in children without first-order belief attribution capacities ($t(18) = 0.17$, adjusted $p = 0.87$, $\eta^2 = 0.07$, 95% *CI* [0.00, 1.00]).

3.3 | Predictors of Intention to Harm

The link between dehumanization and intent to harm varied across children with different levels of theory-of-mind. Dehumanization did not predict harm in cases where participants had only passed the first-order belief attribution tasks ($t(37) = 0.40$, adjusted $p = 0.99$, $\eta^2 < 0.01$, 95% *CI* [0.00, 1.00]) or failed them ($t(17) = -0.43$, adjusted $p = 0.99$, $\eta^2 < 0.01$, 95% *CI* [0.00, 1.00]). However, a link was detected between dehumanization and the intent to harm among children capable of second-order belief attribution ($t(46) = 3.60$, adjusted $p = 0.004$, $\eta^2 = 0.26$, 95% *CI* [0.09, 1.00]) even after controlling for verbal ability ($t(46) = -0.90$, adjusted $p = 0.99$, $\eta^2 = 0.02$, 95% *CI* [0.00, 1.00]),

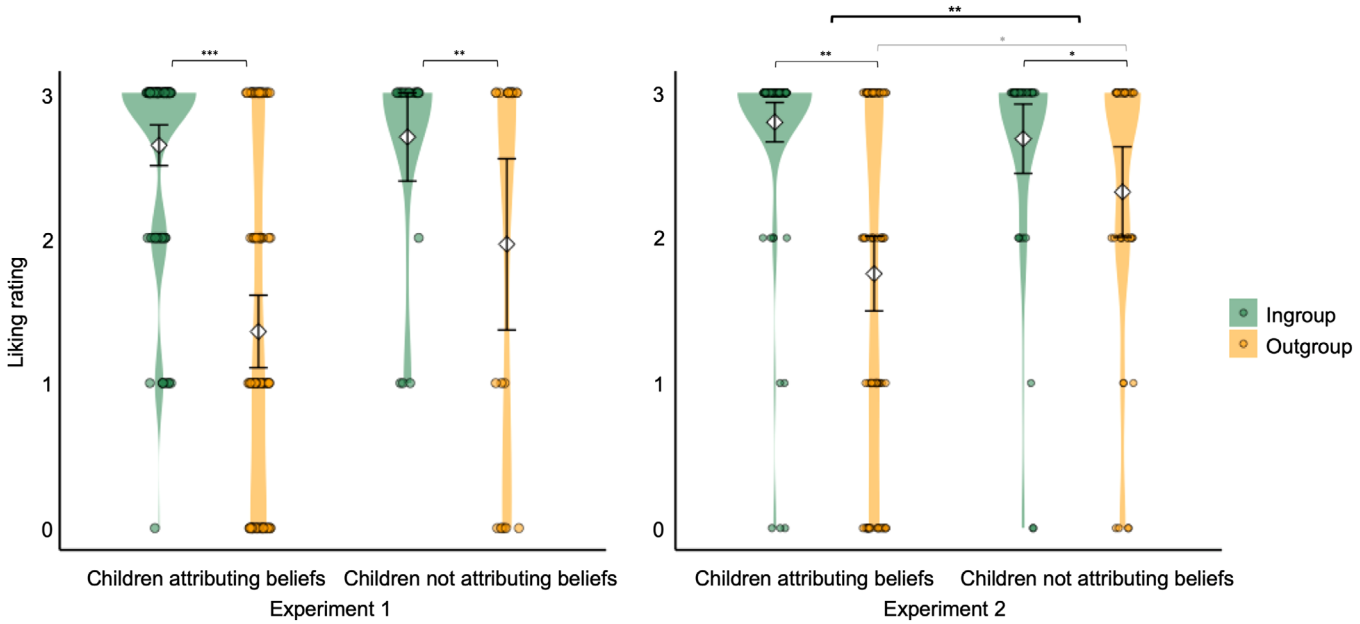

FIGURE 6 | Liking ratings of the ingroup and the outgroup among children who could and could not attribute first-order beliefs. Error bars represent 95% *CI*s. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

age ($t(46) = -0.26$, adjusted $p = 0.99$, $\eta^2 < 0.01$, 95% *CI* [0.00, 1.00]) and gender ($t(46) = 2.03$, adjusted $p = 0.19$, $\eta^2 = 0.10$, 95% *CI* [0.01, 1.00]). In this subsample of children, this predictive effect of dehumanization on harm goes marginally significant ($t(45) = 2.69$, adjusted $p = 0.06$, $\eta^2 = 0.27$, 95% *CI* [0.10, 1.00]) after controlling for intergroup liking ($t(45) = 2.02$, adjusted $p = 0.24$, $\eta^2 = 0.08$, 95% *CI* [0.00, 1.00]). Given the correlation between dehumanization and intergroup liking (r (49) = 0.38, $p = 0.005$) in this subsample, we assessed multicollinearity by calculating the variance inflation factor (VIF). All VIF values were below 5 (range = 1.09–1.27), indicating that multicollinarity is not a concern in this model. On its own, intergroup liking is significantly associated with the intent to harm ($t(109) = 3.59$, adjusted $p = 0.002$, $\eta^2 = 0.09$, 95% *CI* [0.02, 1.00]).

4 | Discussion

Our preliminary results provide initial support for the predicted relationship between the development of theory-of-mind capacities and dehumanization (Hare 2017). Only the children who passed a belief attribution task dehumanized the outgroup more than the ingroup. The more children showed skills on the belief attribution tasks, the stronger they dehumanized the outgroup. They rated the outgroup as less human, rather than shifting their ratings of ingroup humanness. Children with the most advanced belief attribution abilities and strongest dehumanization tendencies seemed most likely to express the intent to harm the outgroup (see also Zhou and Hare 2022).

Our findings also reveal the developmental trajectories of dehumanization in relation to liking. Intergroup difference on liking ratings appeared before the emergence of belief attribution abilities. However, different from previous research that deployed almost the same scales (Mcloughlin et al. 2018), we still found a medium correlation between dehumanization and intergroup liking. The relationship between dehumanization and harm also reduced to marginal significance after controlling for intergroup liking among children who could attribute second-order beliefs. In children with less developed theory-of-mind capacities, dehumanization could not explain preferences for harm.

One limitation of Experiment 1 is the relatively small portion of children showing no belief attribution ($n = 23$). Critically, these children did not differ on the comprehension of humanness measures from children who showed attribution abilities (See the Supporting Information, p. 3). Further, our sensitivity analyses provide additional evidence that our unbalanced comparisons are sufficiently powered (See Tables S1 and S2). Still, a replication focused on the youngest ages in our sample, who are most likely to lack attribution abilities (3–4 years of age), is needed to test the robusteness of our findings.

Since our scales illustrate humanness in an ascending order and children were told they won the game before the humanness ratings begun, it is possible that children rated ingroup and outgroup based on the perception of intergroup hierarchies. Previous work has shown that the endorsement of group-based dominance is associated with blatant dehumanization in both adults and school-age children (Kteily et al. 2015; Zhou and Hare 2022). The same association may appear in preschoolers as well. Notably, children as young as three can infer dominance-subordination relationships (Charafeddine et al. 2015), with a preference for dominance emerging as early as toddlerhood (Thomas et al. 2018), well before theory-of-mind becomes full-blown. However, we did not find dehumanization in children without belief attribution capacities, suggesting that preference for ingroup dominance alone might not influence humanness ratings in the absence of belief understanding. In Experiment 2, we additionally

incorporated an assessment of intergroup hierarchy to further test its role in the development of dehumanization.

Finally, our use of facial stimuli did not match the participants in age or ethnicity. In the introductory session, children demonstrated comprehension of the humanness scale, and their average ingroup ratings were even higher than those reported in a Western sample of 5- and 6-year-olds who used the same facial stimuli to rate humanness (Mcloughlin et al. 2018). However, it is still possible that the scale was not fully aligned with participants' understanding of ingroup humanness. To allow for a qualitative comparison between Experiment 1 and 2, we used the same graphic scale in Experiment 2. However, an additional control test of children's self-evaluation of humanness on the same scale is needed to test how similar they see the illustration of the ingroup to the self.

4.1 | Experiment 2

Here we retest for the proposed link between mentalizing capacities and dehumanization by focusing on 3- to 4-year-olds. In Experiment 1, it was this age group that showed the most variable skills in solving first-order theory-of-mind problems. By focusing on this younger age group we expect a more balanced sample of participants with and without belief attribution capacities.

We also expand the range of belief attribution tasks. Children do not develop an understanding of all categories of beliefs and false beliefs simultaneously with variation in the order in which types of beliefs are understood across development (Wellman and Liu 2004). Therefore we retain our tasks from Experiment 1 that assess explicit false belief and knowledge access while adding a modified "smarties" task designed to test recognition of false belief regarding the contents in a container (Gopnik and Astington 1988). We also examined second-order belief attribution in two of the tasks to retest if more advanced mentalizing skills are related to humanness evaluations.

Regarding measures of humanness, we added a measure to assess participants' self-ratings. We predicted that children would rate themselves as similar to the group identified in the introduction as their own ingroup. We also anticipated that children capable of belief attribution would again rate the outgroup as less human than their ingroup, as well as themselves.

We additionally assessed participants' perception of intergroup hierarchy. This allowed for a supplemental test to determine whether endorsing ingroup dominance is associated with dehumanization in children with different theory-of-mind skills.

5 | Methods

5.1 | Power Analysis

G*Power (Faul et al. 2007) indicated that, with a repeated measures design and an alpha level of .05, a sample size of 54 participants was needed to obtain 80% power to detect the previously observed medium interaction effect ($\eta^2 = 0.07$) between mind attribution and group identity on humanness rating in Experiment 1.

5.2 | Participants

We built an initial participant pool by posting an advertisement through local media, and invited all 3- and 4-year-olds through their parents, excluding any child who had participated in Experiment 1. We successfully recruited 130 children (Mean age = 47.85 months, $SD = 6.05$, range = 36–59 months; 64 boys, 66 girls; All Asians and native Mandarin Chinese speakers). An additional 12 participants were tested but then excluded from all analysis. They either did not complete the testing ($n = 2$), failed the comprehension checks in the belief attribution tasks ($n = 2$) or failed the introductory session of the humanness scales ($n = 8$). Age and gender information was obtained from the parents.

5.3 | Procedure

The procedure was similar to Experiment 1. Participants first finished all three theory of mind tasks. Then, after the minimal intergroup manipulation identical to Experiment 1, participants were introduced to the humanness scale, rated humanness and liking of the ingroup and the outgroup, and reported their intent to harm the outgroup. Additionally, participants rated their own humanness and their perception of a hierarchy between human groups. Below, we outline details of the modified theory-of-mind tasks, the additional humanness self-rating scale, and the intergroup hierarchy assessment.

Theory-of-Mind: We tested first-order theory-of-mind capacity in three tasks, and further tested second-order theory of mind capacity using additional questions that followed up on two of the first-order tasks.

Task 1 tests explicit false-belief. We reused the Sally-Anne test to test first-order belief attribution, and added a follow-up test to assess second-order belief attribution. In this follow-up, E1 explained that when Fangfang moved the ball, Lili peeked at what was happening. E1 then asked the participant a comprehension question, "Where will Fangfang actually look for the ball when she is back?". Then E1 assessed the participant's capacity for second-order belief attribution by asking, "When Fangfang comes back, where Lili will think Fangfang will look for the ball?". The participant was also asked to explain their reasoning. Responses were coded as correct only if they 1) correctly answered the first-order question, the comprehension check, and the second-order question, and 2) explained that Fangfang did not know Lili had peeked.

Task 2 is identical to the surprise task of Experiment 1, testing participants' performances on first- and second-order knowledge-ignorance reasoning and detection of the hidden emotion.

Task 3 adapts the "Smarties" contents false belief test (Gopnik and Astington 1988). E1 first showed an opaque box of crayons to both E2 and the participant, ensuring that everyone understood the box contained crayons. After confirming this, E2 left the

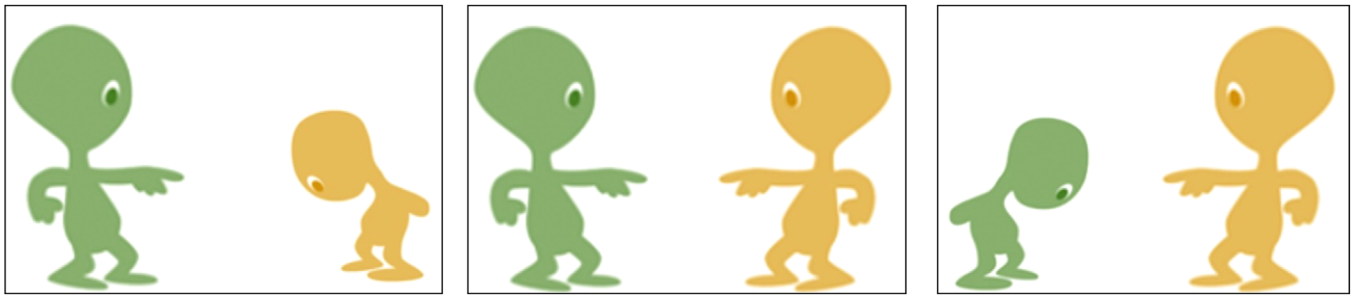

FIGURE 7 | Illustrations depicting different intergroup relations used in the hierarchy perception task.

testing room. E1 then removed the crayons, hid them, and replaced them with four pencils inside the box. Then, E1 asked the participant a comprehension question: "What is inside the box now?". This was followed by the first-order belief attribution question, "When E2 comes back, what will they think is inside the box?". Participant responses were coded as correct if they answered "crayons".

Consistent with Study 1, theory-of-mind questions were coded using a binary pass/fail scale (0 = incorrect, 1 = correct), A participant was considered as having first-order theory-of-mind if they passed at least one first-order question, and capable of second-order belief attribution if they passed at least one second-order question.

To further understand if flexibility in theory-of-mind application matters, we also calculated if participants succeeded in different types of tasks through a cumulative first-order score. This score reflected the number of first-order task types passed Responses to all theory-of-mind questions were coded as either "0" (*incorrect*) or "1" (*correct*). This coding scheme allowed for measuring the appearance of first-order belief attribution ("0" = *failed all first-order questions*, "1" = *passed at least one first-order question), and* the appearance of second-order belief attribution *("*0" = *failed*, *"1" = passed at least one second-order question*). We also measured if participants flexibly attributed different categories of beliefs by accumulating responses to first-order questions across the three tasks (from "o" = failed all questions to "3" = passed all first-order questions).

Identical to Study 1, we coded participants as understanding hidden emotions if they passed the emotion detection question.

Self Humanness: Each participant rated how much they fit the definition of humanness using the four-point scales that they used to rate the ingroup and outgroup humanness.

Hierarchy Perception: We adopted a paradigm that was designed to test if preschoolers recognize and value dominance (Charafeddine et al. 2019). Each participant was presented three figures together, in which two characters matched the colors of the ingroup and the outgroup respectively (Figure 7). The participant was asked "Which figure do you think best represent the relationship between your team and your opponents, the yellow/green team?". Responses were coded as *"dominant"* (character with ingroup color extending and showing power), *"equal"* (two characters sizing equally), and *"subordinate"* (character with ingroup color shrinking).

5.4 | Analysis

Using ANCOVA, we examined the interaction between theory-of-mind capacities and the evaluated group's identity (ingroup/outgroup/self) on humanness ratings. We also tested whether the number of belief categories successfully attributed by participants is associated with their humanness ratings for the different groups. Additionally, we explored the interaction between hierarchy perception and the evaluated group's identity on humanness ratings in children with and without theory-of-mind capacities. Hierarchy perception was treated as a categorical variable.

The plans for comparing intergroup liking ratings and analyzing the relationship between dehumanization and liking were identical to Experiment 1. Since only 15 three- and four-year-olds demonstrated second-order belief attribution, the analysis of the relationship between dehumanization and outgroup harm in this subsample was underpowered and therefore moved to the Supporting Information (Table S4). For the main analysis, we calculated linear regressions to examine the predictors of harm in children with and without first-order belief attribution capacities.

We included age by months and gender as demographic covariates in all analyses. All tests were two-tailed. Significance values were adjusted using Holm-Bonferroni method.

6 | Results

Among the participants, 89 (41 females, 48 males; Mean age = 48.88 months, *SD* = 6.04 months, range = 36–59 months) passed at least one belief attribution question, demonstrating some theory-of-mind capacities. The other 41 participants (25 females, 16 males; Mean age = 45.63 months, *SD* = 5.53 months, range = 36–58 months) did not show any evidence of belief attribution. Figure 3B illustrates participants' belief attribution capacities by age.

6.1 | The Relationship Between Dehumanization and Mind Attribution

Consistent with the results of Experiment 1, on humanness ratings, ANCOVA revealed a significant interaction between first-order belief attribution and the evaluated group's identity ($F(2,251) = 12.22$, adjusted $p < 0.001$, $\eta^2 = 0.09$, 95% *CI* [.04, 1.00]; See Figure 4). Post hoc comparisons showed that participants with belief attribution capacities gave significantly lower

humanness ratings to the outgroup (Mean = 1.73, 95% *CI* [1.47, 1.99]) compared to the ingroup (Mean = 2.54, 95% *CI* [2.33, 2.75]; $t(251) = -5.25$, adjusted $p < 0.001$, Cohen's $d = -0.71$, 95% *CI* [−1.02, −0.41]), indicating evidence of outgroup dehumanization. In contrast, participants without belief attribution capacities did not show significant differences in humanness ratings between the outgroup (Mean = 2.46, 95% *CI* [2.16, 2.76]) and ingroup (Mean = 2.05, 95% *CI* [1.64, 2.46]; $t(251) = 1.40$, $p = 0.34$, Cohen's $d = 0.36$, 95% *CI* [−0.07, 0.80]). Similarly, only children with belief attribution rated the outgroup as less human than themselves (self: Mean = 2.56, 95% *CI* [2.29, 3.18]; $t(251) = -6.47$, adjusted $p < 0.001$, Cohen's $d = -0.90$, 95% *CI* [−1.21, −0.59]), whereas children without belief attribution did not show such a difference (self: Mean = 2.68, 95% *CI* [2.48, 2.87]; $t(251) = -0.59$, $p = 0.82$, Cohen's $d = -0.11$, 95% *CI* [−0.54, 0.33]). Participants did not differentiate between ingroup and self humanness ratings regardless of having belief attribution ($t(251) = -1.25$, $p = 0.42$, Cohen's $d = -0.16$, 95% *CI* [−0.45, 0.14]) or not ($t(251) = -1.99$, $p = 0.12$, Cohen's $d = -0.46$, 95% *CI* [−0.90, −0.02]). When directly comparing to children without belief attribution, children capable of belief attribution rated the outgroup significantly less human ($t(357) = -3.37$, adjusted $p < 0.001$, Cohen's $d = -0.63$, 95% *CI* [−1.01, −0.25]). We did not see difference in their ratings of ingroup humanness ($t(357) = 1.81$, $p = 0.07$, Cohen's $d = 0.44$, 95% *CI* [0.07, 0.82]) or self humanness ($t(357) = 0.59$, $p = 0.56$, Cohen's $d = 0.14$, 95% *CI* [−0.23, 0.51]). On covariates, we did not find age ($F(2,251) = 2.64$, $p = 0.07$, $\eta^2 = 0.02$, 95% *CI* [0.00, 1.00]; See the distribution of humanness ratings by age in Figure S2) or gender ($F(2,251) = 2.19$, $p = 0.11$, $\eta^2 = 0.02$, 95% *CI* [0.00, 1.00]) interacting with group identity to influence humanness ratings.

The number of belief categories that children successfully attributed is also associated with their humanness ratings of the different groups ($F(2, 251) = 10.11$, adjusted $p < 0.001$, $\eta^2 = 0.07$, 95% *CI* [0.03, 1.00]). As children succeeded with more first-order tasks, their rating of outgroup humanness decreased ($t(359) = -3.07$, adjusted $p = 0.007$, $\beta = -0.31$, 95% *CI* [−0.56, −0.07]) while the ratings of the ingroup ($t(359) = 1.16$, $p = 0.24$, $\beta = 0.12$, 95% *CI* [−0.13, 0.36]) and self ($t(359) = -0.09$, $p = 0.93$, $\beta = 0.01$, 95% *CI* [−0.25, 0.24]) were not significantly different from reports of children without theory-of-mind capacities. Post hoc analyses indicated a significant difference in the magnitude of the ingroup–outgroup rating differences ($t(251) = -3.25$, adjusted $p = 0.004$, $\beta = -0.43$, 95% *CI* [−0.75, −0.12]). We did not find any effect of age $F(2,251) = 1.42$, adjusted $p = 0.36$, $\eta^2 = 0.01$, 95% *CI* [0.00, 1.00]) or gender $F(2,251) = 2.12$, adjusted $p = 0.65$, $\eta^2 = 0.02$, 95% *CI* [0.00, 1.00]).

Regarding the role of second-order belief attribution, we did not find an interaction effect with the evaluated group's identity on humanness ratings ($F(2,169) = 0.89$, adjusted $p = 0.83$, $\eta^2 = 0.01$, 95%*CI* [0.00, 1.00]; See Figure 5), and only found a significant main effect of group identity ($F(2,169) = 23.42$, adjusted $p < 0.001$, $\eta^2 = 0.22$, 95%*CI* [0.13, 1.00]). In participants who passed second-order belief attribution ($n = 15$), the outgroup (Mean = 1.40, 95% *CI* [0.71, 2.09]) was rated as less human than the ingroup (Mean = 2.60, 95% *CI* [2.10, 3.10]; $t(169) = -2.84$, adjusted $p = 0.01$, Cohen's $d = -1.10$, 95%*CI* [−1.86, −0.32]) and oneself (Mean = 2.73, 95% *CI* [2.29, 3.18]; $t(169) = -2.91$, adjusted $p = 0.01$, Cohen's $d = -1.28$, 95%*CI* [−2.06, −0.48]). The ratings of the ingroup and the self did not differ ($t(169) = -0.07$, adjusted $p = 0.99$, Cohen's $d = -0.16$, 95%*CI* [−0.87, 0.56]). These findings were similar to ratings by participants who only passed first-order belief attribution ($n = 74$; Outgroup: Mean = 1.80, 95% *CI* [1.51, 2.09]; Ingroup: Mean = 2.53, 95% *CI* [2.29, 2.76]; Self: Mean = 2.67, 95% *CI* [2.48, 2.87]; Outgroup—Ingroup: $t(169) = -4.44$, adjusted $p < 0.001$, Cohen's $d = -0.47$, 95%*CI* [−0.77, −0.17]; Outgroup—Self: $t(169) = -5.62$, $p < 0.001$, Cohen's $d = -0.82$, 95%*CI* [−1.12, −0.51]; Ingroup—Slef: $t(169) = -1.20$, $p = 0.45$, Cohen's $d = -0.11$, 95%*CI* [−0.41, 0.18]). We did not find any effect of age $F(2,169) = 0.82$, adjusted $p = 0.83$, $\eta^2 < 0.01$, 95% *CI* [0.00, 1.00]) or gender $F(2,169) = 2.36$, adjusted $p = 0.29$, $\eta^2 = 0.03$, 95% *CI* [0.00, 1.00]).

Finally, consistent with the results of Experiment 1, we found no link between hidden-emotion detection and intergroup differences in humanness ratings ($F(2, 251) = 3.47$, adjusted $p = 0.07$, $\eta^2 = 0.03$, 95% *CI* [0.00, 1.00]; See results of post hoc analyses in Tables S3).

6.2 | The Role of Hierarchy Perception

Regarding the role of hierarchy perception, we did not find an interaction with the evaluated group's identity on humanness ratings in children without belief attribution ($F(2, 84) = 0.12$, $p = 0.89$, $\eta^2 < 0.01$, 95% *CI* [0.00, 1.00]). This did not change when children developed belief attribution ($F(2, 36) = 0.21$, $p = 0.81$, $\eta^2 = 0.01$, 95% *CI* [0.00, 1.00]).

6.3 | Distinguishing Between Dehumanization and Intergroup Liking

On liking ratings, we detected an interaction effect between belief attribution and the evaluated group's identity ($F(1, 126) = 9.82$, adjusted $p = 0.006$, $\eta^2 = 0.07$, 95% *CI* [0.02, 1.00]; See Figure 6). Post hoc comparisons indicated that liking ratings are significantly lower for outgroup than ingroup in both children capable of first-order belief understanding (Outgroup: Mean = 1.75, *95% CI* [1.50, 2.01]; Ingroup: Mean = 2.80, *95% CI* [2.66, 2.93]; $t(126) = -8.17$, adjusted $p < 0.001$, Cohen's $d = -1.07$, 95% *CI* [−1.39, −0.76]) and those who were incapable (Outgroup: Mean = 2.32, *95% CI* [2.01, 2.63]; Ingroup: Mean = 2.68, *95% CI* [2.44, 2.92]; $t(126) = -2.42$, adjusted $p = 0.02$, Cohen's $d = -0.42$, 95% *CI* [−0.85, 0.02]). However, children with first-order belief understanding liked the outgroup less when directly comparing to children showing no evidence of this ability ($t(236) = -2.49$, adjusted $p = 0.01$, Cohen's $d = -0.49$, 95% *CI* [−0.86, −0.11]), while their ratings for the ingroup did not differ ($t(236) = 0.54$, $p = 0.59$, Cohen's $d = 0.17$, 95% *CI* [−0.20, 0.54]). On covariates, we found a significant interaction of group identity with age ($F(1, 126) = 5.30$, adjusted $p = 0.046$, $\eta^2 = 0.04$, 95% *CI* [0.00, 1.00]), older children being less likely to like the outgroup than the ingroup ($t(126) = -2.31$, adjusted $p = 0.02$, $\beta = -0.04$, 95% CI [−0.07, −0.01]). But we did not see an effect of gender ($F(1, 126) = 0.12$, $p = 0.73$, $\eta^2 < 0.01$, 95% *CI* [0.00, 1.00]).

In contrast to Experiment 1, we did not find a significant association between intergroup liking and dehumanization in children who understood first-order belief ($t(85) = 2.30$, adjusted $p = 0.12$, $\eta^2 = 0.07$, 95% *CI* [0.01, 1.00]) after controlling for age ($t(85) = 0.81$, adjusted $p = 0.42$, $\eta^2 < 0.01$, 95% *CI* [0.00, 1.00])

and gender ($t(85) = 0.15$, $p = 0.88$, $\eta^2 < 0.01$, 95% *CI* [0.00, 1.00]). This association was absent in children without first-order belief understanding ($t(37) = -0.47$, $p = 0.64$, $\eta^2 < 0.01$, 95% *CI* [0.00, 1.00]), either.

6.4 | Predictors of Harm

As in Experiment 1, dehumanization did not show any effect in predicting outgroup harm in either children who only developed first-order belief attribution ($t(70) = -0.04$, $p = 0.97$, $\eta^2 < 0.01$, 95% *CI* [0.00, 1.00]) or those who did not ($t(37) = 1.05$, $p = 0.30$, $\eta^2 < 0.01$, 95% *CI* [0.00, 1.00]). The intergroup difference in liking ratings also did not show a significant effect in explaining the intent to harm the outgroup ($t(126) = 2.44$, adjusted $p = 0.06$, $\eta^2 = 0.05$, 95% *CI* 0.01, 1.00]). The participants' age or gender again did not influence any of these evaluations (see full results in Table S4).

7 | Discussion

Experiment 2 replicates the findings of Experiment 1 regarding the relationship between theory-of-mind development and outgroup dehumanization. Specifically, only children with belief attribution capacities rated the outgroup less human than the ingroup. Consistent with the findings of Experiment 1, as first-order belief attribution capacities emerged and matured, children rated the outgroup as less human. In contrast, children rated themselves and their ingroup consistently high on the same humanness scale regardless of belief attribution performance. When children developed second-order belief attribution capacities, they dehumanized with a larger effect size. Nevertheless, unlike Experiment 1, we did not find a significant increase of dehumanization in children with second-order belief attribution, but this is likely due to the relatively small subsample of three- and four-year-olds capable of this form of attribution ($n = 15$). We also found no evidence that the perception of intergroup hierarchy impacted our participants' humanness ratings.

Our findings also further underscored the difference between the developmental trajectories of dehumanization and liking. Children consistently liked the outgroup less than the ingroup, and this pattern again appeared before the emergence of belief attribution capacities. In contrast to Experiment 1, we did not find an association between dehumanization and intergroup liking, but similar to the shift of outgroup humanness ratings, children liked the outgroup even less after they understood beliefs. These differences are observed after controlling for age and gender.

Finally, neither dehumanization or the lack of liking explained the intent to cause outgroup harm. While a significant effect of dehumanization was found to predict harm in participants with second-order belief attribution in Experiment 1, given our focus on a younger sample in this experiment, the relatively small number of 3–4 year olds showing second-order capacities did not provide sufficient power to test for that relationship.

7.1 | General Discussion

Across two experiments, we found supporting evidence for the relationship between the development of theory-of-mind and dehumanization (Hare 2017; Hare and Woods 2020). Results consistently showed that only the children with belief attribution capacities considered the outgroup as less human. Children with more advanced belief attribution, indicated by robust performance across diverse tasks and the development of second-order reasoning, also demonstrated stronger dehumanization, reflected in lower humanness ratings for the outgroup. In contrast, their self and ingroup ratings remained consistent regardless of belief attribution performance. While previous findings indicate that children with more advanced theory-of-mind are more likely to incorporate others' mental states into moral reasoning (Fu et al. 2014; Heron-Delaney et al. 2011), our results imply this advanced capacity may be applied selectively and further privilege ingroup over outgroup members in attributions of humanness. Moreover, children with second-order theory-of-mind skills may be more attuned to social expectations for outgroup negativity and act accordingly (Abrams et al. 2009; Rhodes 2012).

Our results provide little evidence the observe patterns are due to a more generalized socio-cognitive process (i.e. as assessed through verbal ability) or ingroup liking differences. The difference between children's ingroup and outgroup humanness ratings was not associated with verbal ability abilities or age, and all the effects remain significant after controlling for these covariates. Thus, the significant relationships are likely not signaling a general developmental pattern across many co-emergent social skills. Although research in adults often indicate a correlation between intergroup liking and a tendency to dehumanize (Costello and Hodson 2014; Kteily et al. 2015), this association seems fragile in young children, appearing only in the sample with an older average age (Experiment 1). While both liking-disliking and dehumanization increase with development, the two appear developmentally distinguishable based, in part, on the relationship of dehumanization to the emergence of belief attribution abilities. In addition, different from what was found in adults and school-age children (Kteily et al. 2015; Zhou and Hare 2022), assumptions about ingroup dominance over the outgroup did not contribute to the strength of dehumanization in our preschool participants. This makes it difficult to interpret participants dehumanizing tendencies as a by-product of a general desire for ingroup dominance.

The link between dehumanization and antisocial behaviors is well-established in previous research on adults and school-age children (Kteily and Bruneau 2017; van Noorden et al. 2014; Zhou and Hare 2022). However, we only found dehumanization in association with harmful intent among preschoolers with the most advanced belief attribution in Experiment 1. Notably, this association dropped to marginal significance after controlling for liking, suggesting that while dehumanization and liking are likely distinct constructs (Mcloughlin et al. 2018), their effects on harmful intent interact. In contrast, children with more limited attributional abilities might not fully understand how their choices could cause harm to a recipient, or minimal group manipulations may not provide strong enough motivation for the expression of harmful intent (Rhodes 2012). Together, our

findings suggest the possibility that dehumanization and the intent to harm are not as tightly linked in early childhood. While replications with larger samples are needed, this preliminary evidence points to a potential window before dehumanization becomes tied to harmful acts, Interventions during this window that guide children's fledgling theory-of-mind toward more inclusive moral concern may help disrupt the consolidation of dehumanization and reduce the likelihood that it facilitates aggression later in life.

It is noteworthy that not all participants in our experiments rated their ingroup or themselves as fully human. First, even groups of adults typically do not rate their own ingroup as fully human (Kteily and Bruneau 2017). Second, this pattern is unlikely to be driven by ethnic differences between the children and the faces used in the scale, as similar findings have been reported among 6-year-old children in the UK who first used this scale and should be more familiar with the faces it depicts. Their average ratings of their ingroup was even lower on the same facial humanness continuum (Mcloughlin et al. 2018). Another possible explanation is that children are often viewed as human becomings and not yet fully human in the sociocultural sense (Hagá, 2023). Our findings, alongside those of Mcloughlin et al. (2018), imply that children across different cultural contexts may also internalize this perspective.

As an initial attempt to test the association between theory-of-mind development and perception of humanness in preschoolers, the current experiments also have a number of potential limitations that suggest directions for future research. First, we find no relationship between our supplemental emotional sensitivity measure and dehumanization. We suspect this is due to the binary nature of our single measure. It is possible a wider range of direct emotion detection tasks, like the "Reading the Mind in the Eyes Test" (Baron-Cohen et al. 2001), will reveal a relationship between emotion attribution abilities and dehumanization[1]. It will also be essential to explore the role of individual variability in intentional, emotional and perceptual attribution. We strongly suspect other attribution measures will be related to dehumanization as well (Harris 2017). Second, our experiment was only designed to test for a blatant and broad form of dehumanization, grounded in a developmentally emerging concept of humanness. It is noteworthy that humanness can be multidimensional and consist of components beyond mind (Fincher et al. 2018; Zhou et al. 2024). Although our comprehension check supports participants understanding of the ordinal nature of the graphic metaphors, it remains unknown if their conception of humanity was based on physical human-likeness, metaphor or on more abstract qualities. While a paucity of nonverbal measures remains a major constraint, a priority will be testing the developmental relationship of mentalizing and multiple forms of dehumanization. For example, future experiments can be designed to test how children conceive of the outgroup as either having antisocial traits but lacking positive human qualities, or somehow denying humanness more globally (Over 2021). It may also be possible to use more subtle measures to better differentiate if children are denying emotional capacity, agency or other components of humanity (e.g., Bain et al. 2009). A third limitation concerns the specific context of our research. The competitive setting, which was designed to maximize the likelihood of observing dehumanization (Zhou and Hare 2022), limits the generalizability of our findings to other social situations. Future work should test this relationship in more diverse contexts, ideally using longitudinal designs to delineate the causal developmental pathway that our cross-sectional data cannot establish. Since our participants represent a relatively homogenous sample being from a single nationality, it is also theoretically urgent to explore whether the same link between theory-of-mind development and dehumanization occur in other cultural contexts using multiple cross-cultural measures. There is also the possibility that the fixed order of task presentation influenced participants responses, and a different order of presentation may yield different results. A particularly important test would be to vary whether the theory of mind task precedes the intergroup task, as mentalizing before evaluating groups may increase outgroup dehumanization rather than disliking.

Finally, our preschool participants have finite attentional resources. Our design prioritized conciseness, but leaves us unable to completely rule out the possibility that the relationship we observed between dehumanization and theory-of-mind is explained by a third unmeasured trait. One plausible candidate might be executive function which is thought to be critical to human unique mentalizing and group decision making (Hare 2017). Participants who failed the theory-of-mind tasks might simply lack the level of executive function needed to attribute false-beliefs or manifest bias on the dehumanization scale (Devine and Hughes 2014;but see Wang et al. 2016). However, two findings challenge this interpretation. First, participants in both experiments without false-belief attribution still exhibited a robust ingroup preference in their liking judgments. This confirms they possessed enough executive capacity to understand the scales, and express an ingroup social bias. Second, if executive function alone drives the observed link between theory-of-mind and dehumanization, then children with more advanced theory-of-mind (and presumably greater executive function, Baker et al. 2021) should be less prone to dehumanization, as they could have more cognitive skills to overcome in-group bias (Hoyo et al. 2019; Yu et al. 2016). However, we found that advanced theory-of-mind was associated with stronger dehumanization. Moreover, while executive function is know to increase with age (Devine and Hughes 2014), we did not find a relationship between age and dehumanization that should exist if executive function was driving its emergence. This suggests that while executive function facilitates complex social judgments, the emergence of dehumanization is also associated with attributional abilities. Future experiments can include measurse of executive function to further test this interpretation.

Despite the potential limitations, the current experiments make a novel contribution by examining dehumanization in 3- to 6-year-olds and by including a relatively rare non-WEIRD sample. Together, these findings offer initial evidence for a potential developmental pattern for dehumanization. At birth, infants distinguish humans from nonhuman animals (Axelsson et al. 2018; Haan et al. 2002; Pauen 2000). Soon after, and well before the age of three, in-group favoritism emerges (Over 2018). As theory-of-mind develops, children next integrate mental life into their representation of humanness. They eventually see others as having a mental life that can differ from their own, and apply this capacity selectively, attributing more mental states to ingroup than outgroup members (McLoughlin and Over 2017). At this point, dehumanization emerges. While children begin showing

human-unique forms of cooperation and prosociality that rely on mental attribution, they also develop the potential to see a group challenging their own as less than fully human and become more accepting of harm toward them.

Overall, our results support the hypothesis that the ability to attribute or deny mental states is associated with the attribution and denial of humanness according to group membership. This overall pattern may be related to the observed link in adults between dehumanization and reduced empathic responses (Harris and Fiske 2006; Jack et al. 2013; Simon and Gutsell 2021), and will contribute to debates on the precise psychology of dehumanization (Kteily and Landry 2022; Rai et al. 2018). Importantly, our findings highlight that theory-of-mind is not inherently prosocial but functions flexibly, modulated by group membership. If correct, there is urgent need for strategies and interventions that channel theory-of-mind toward broader moral inclusion, for instance, by fostering awareness of outgroup members' mental lives and establishing shared identities and interdependence through cooperative learning and early intergroup contact (Bruneau et al. 2021; Wright et al. 2014). The most powerful interventions will generalize to previously unencountered outgroups (Lolliot et al. 2013) and will increase the recognition of humanness and moral concern for all humans from the youngest ages into adulthood.

Author Contributions

All authors contributed to the study design. W.Z. performed testing, data collection, data analysis and interpretation under the supervision of B.H. W.Z. and B. H. wrote the manuscript. All authors approved the final version of the manuscript for submission.

Acknowledgment

We thank all the participants for taking part in the research, the staff of the kindergarten for continuous support, and the research assistants who helped with data collection. We also thank Drs. Michael Tomasello and Tamar Kushnir for helpful comments on an earlier version of this manuscript. Funding for this project was provided by Duke Kunshan University. The work of BH is supported in part by grant 62280 from the John Templeton Foundation.

Ethics Statement

Ethics approval for this study was granted by Duke Kunshan University Campus Institutional Review Board (Protocol No. 2022WZ122). All child participants' parents were provided an opt-out parental permission form. Verbal assent from each child participant was obtained prior to the experiment.

Conflicts of Interest

All authors declare no conflicts of interest.

Artificial Intelligence

No artificial intelligence assisted technologies were used in this research or the creation of this article.

Data Availability Statement

Preregistration: For both experiments, the materials, data and analysis scripts are publicly available (https://osf.io/bndaz/). Experiment 1 was not preregistered. We preregistered the hypotheses, methods and analysis plan of Experiment 2 prior to data collection (https://osf.io/e82yt).

Endnotes

[1] It also might be at the emotional level that liking most heavily modulates whether individuals see others as possessing a full human mind and even what kind of emotional states others are attributed (Enock et al. 2021). Alternatively, it may be that sensitivity to other's emotional states has been overestimated in its importance during dehumanization (Bloom 2017).

References

Abrams, D., A. Rutland, J. Pelletier, and J. M. Ferrell. 2009. "Children's Group Nous: Understanding and Applying Peer Exclusion Within and Between Groups." *Child Development* 80, no. 1: 224–243.

Axelsson, E. L., D. G. Moore, E. M. Murphy, J. E. Goodwin, and B. R. Clifford. 2018. "The Role of Bodies in Infants' Categorical Representations of Humans and Non-Human Animals." *Infant and Child Development* 27, no. 6: e2112.

Bain, P., J. Park, C. Kwok, and N. Haslam. 2009. "Attributing Human Uniqueness and Human Nature to Cultural Groups: Distinct Forms of Subtle Dehumanization." *Group Processes & Intergroup Relations* 12, no. 6: 789–805.

Baker, E. R., A. P. D'Esterre, and J. P. Weaver. 2021. "Executive Function and Theory of Mind in Explaining Young Children's Moral Reasoning: a Test of the Hierarchical Competing Systems Model." *Cognitive Development* 58: 101035.

Baron-Cohen, S., A. M. Leslie, and U. Frith. 1985. "Does the Autistic Child Have a "Theory of Mind"?" *Cognition* 21, no. 1: 37–46.

Baron-Cohen, S., S. Wheelwright, J. Hill, Y. Raste, and I. Plumb. 2001. "The "Reading the Mind in the Eyes" Test Revised Version: A Study With Normal Adults, and Adults With Asperger Syndrome or High-Functioning Autism." *The Journal of Child Psychology and Psychiatry and Allied Disciplines* 42, no. 2: 241–251.

Bloom, P. 2017. "The Root of all Cruelty." *The New Yorker* 27: 74–77.

Brewer, M. B. 2001. "Ingroup Identification and Intergroup Conflict." *Social Identity, Intergroup Conflict, and Conflict Reduction* 3: 17–41.

Brewer, M. B. 2007. "The Importance of Being We: Human Nature and Intergroup Relations." *American Psychologist* 62, no. 8: 728.

Bruneau, E., B. Hameiri, S. L. Moore-Berg, and N. Kteily. 2021. "Intergroup Contact Reduces Dehumanization and Meta-Dehumanization: Cross-Sectional, Longitudinal, and Quasi-Experimental Evidence From 16 Samples in Five Countries." *Personality and Social Psychology Bulletin* 47, no. 6: 906–920.

Bruneau, E., H. Szekeres, N. S. Kteily, L. R. Tropp, and A. Kende. 2020. "Beyond Dislike: Blatant Dehumanization Predicts Teacher Discrimination." *Group Processes & Intergroup Relations* 23, no. 4: 560–577.

Butterfill, S. A., and I. A. Apperly. 2013. "How to Construct a Minimal Theory of Mind." *Mind & Language* 28, no. 5: 606–637.

Cameron, J. A., J. M. Alvarez, D. N. Ruble, and A. J. Fuligni. 2003. "Children's Lay Theories About Ingroups and Outgroups: Reconceptualizing Research on Prejudice." In *Lay Theories and Their Role in the Perception of Social Groups*, 118–128. Psychology Press.

Čehajić, S., R. Brown, and R. González. 2009. "What Do I Care? Perceived Ingroup Responsibility and Dehumanization as Predictors of Empathy Felt for the Victim Group." *Group Processes & Intergroup Relations* 12, no. 6: 715–729.

Charafeddine, R., H. Mercier, F. Clément, et al. 2015. "How Preschoolers Use Cues of Dominance to Make Sense of Their Social Environment." *Journal of Cognition and Development* 16, no. 4: 587–607.

Charafeddine, R., H. Mercier, T. Yamada, et al. 2019. "Cross-Cultural Differences in the Valuing of Dominance by Young Children." *Journal of Cognition and Culture* 19, no. 3–4: 256–272.

Costello, K., and G. Hodson. 2014. "Explaining Dehumanization Among Children: The Interspecies Model of Prejudice." *British Journal of Social Psychology* 53, no. 1: 175–197.

Devine, R. T., and C. Hughes. 2014. "Relations Between False Belief Understanding and Executive Function in Early Childhood: A Meta-Analysis." *Child Development* 85, no. 5: 1777–1794.

Dodell-Feder, D., S. H. Lincoln, J. P. Coulson, and C. I. Hooker. 2013. "Using Fiction to Assess Mental State Understanding: A New Task for Assessing Theory of Mind in Adults." *PLoS ONE* 8, no. 11: e81279.

Enock, F. E., S. P. Tipper, and H. Over. 2021. "Intergroup Preference, Not Dehumanization, Explains Social Biases in Emotion Attribution." *Cognition* 216: 104865.

Epley, N., A. Waytz, and J. T. Cacioppo. 2007. "On Seeing Human: A Three-Factor Theory of Anthropomorphism." *Psychological Review* 114, no. 4: 864.

Farroni, T., G. Csibra, F. Simion, and M. H. Johnson. 2002. "Eye Contact Detection in Humans From Birth." *Proceedings of the National Academy of Sciences* 99, no. 14: 9602–9605.

Faul, F., E. Erdfelder, A.-G. Lang, and A. Buchner. 2007. "G* Power 3: A Flexible Statistical Power Analysis Program for the Social, Behavioral, and Biomedical Sciences." *Behavior Research Methods* 39, no. 2: 175–191.

Fincher, K. M., N. S. Kteily, and E. G. Bruneau. 2018. "Our Humanity Contains Multitudes: Dehumanization Is More Than Overlooking Mental Capacities." *Proceedings of the National Academy of Sciences* 115, no. 15: E3329–E3330.

Fiske, S. T. 2013. "Varieties of (de) Humanization: Divided by Competition and Status." In *Objectification and (de) Humanization: 60th Nebraska Symposium on Motivation*, 53–71. Springer.

Fu, G., W. S. Xiao, M. Killen, and K. Lee. 2014. "Moral Judgment and Its Relation to Second-Order Theory of Mind." *Developmental Psychology* 50, no. 8: 2085.

Fuhs, M. W., and J. D. Day. 2011. "Verbal Ability and Executive Functioning Development in Preschoolers at Head Start." *Developmental Psychology* 47, no. 2: 404.

Goldman, E. J., and D. Poulin-Dubois. 2024. "Children's Anthropomorphism of Inanimate Agents." *Wiley Interdisciplinary Reviews: Cognitive Science* 15, no. 4: e1676.

Goldstein, M. H., J. A. Schwade, and M. H. Bornstein. 2009. "The Value of Vocalizing: Five-Month-Old Infants Associate Their Own Noncry Vocalizations With Responses From Caregivers." *Child Development* 80, no. 3: 636–644.

Gopnik, A., and J. W. Astington. 1988. "Children's Understanding of Representational Change and Its Relation to the Understanding of False Belief and the Appearance-Reality Distinction." *Child Development* 59, no. 1: 26–37.

Gray, H. M., K. Gray, and D. M. Wegner. 2007. "Dimensions of Mind Perception." *Science* 315, no. 5812: 619–619.

Gray, K., L. Young, and A. Waytz. 2012. "Mind Perception Is the Essence of Morality." *Psychological Inquiry* 23, no. 2: 101–124.

Grossmann, T. 2017. "The Eyes as Windows Into Other Minds: An Integrative Perspective." *Perspectives on Psychological Science* 12, no. 1: 107–121.

Haan, M., O. Pascalis, and M. H. Johnson. 2002. "Specialization of Neural Mechanisms Underlying Face Recognition in Human Infants." *Journal of Cognitive Neuroscience* 14, no. 2: 199–209.

Hackel, L. M., C. E. Looser, and J. J. Van Bavel. 2014. "Group Membership Alters the Threshold for Mind Perception: The Role of Social Identity, Collective Identification, and Intergroup Threat." *Journal of Experimental Social Psychology* 52: 15–23.

Hare, B. 2011. "From Hominoid to Hominid Mind: What Changed and Why?" *Annual Review of Anthropology* 40: 293–309.

Hare, B. 2017. "Survival of the Friendliest: Homo Sapiens Evolved *via* Selection for Prosociality." *Annual Review of Psychology* 68: 155–186.

Hare, B., J. Call, B. Agnetta, and M. Tomasello. 2000. "Chimpanzees Know What Conspecifics Do and Do Not See." *Animal Behaviour* 59, no. 4: 771–785.

Hare, B., J. Call, and M. Tomasello. 2001. "Do Chimpanzees Know What Conspecifics Know?" *Animal Behaviour* 61, no. 1: 139–151.

Hare, B., and V. Woods. 2020. *Survival of the Friendliest: Understanding Our Origins and Rediscovering Our Common Humanity*. Random House.

Harris, L. T. 2017. *Invisible Mind: Flexible Social Cognition and Dehumanization*. MIT Press.

Harris, L. T., and S. T. Fiske. 2006. "Dehumanizing the Lowest of the Low: Neuroimaging Responses to Extreme Out-Groups." *Psychological Science* 17, no. 10: 847–853.

Harris, L. T., and S. T. Fiske. 2011. "Perceiving Humanity or Not: A Social Neuroscience Approach to Dehumanized perception." In *Social Neuroscience: Toward Understanding the Underpinnings of the Social Mind*, 123–134. Oxford University Press.

Haslam, N. 2006. "Dehumanization: An Integrative Review." *Personality and Social Psychology Review* 10, no. 3: 252–264.

Haslam, N., B. Bastian, S. Laham, and S. Loughnan. 2012. "Humanness, Dehumanization, and Moral Psychology." In *The Social Psychology of Morality: Exploring the Causes of Good and Evil*, edited by M. Mikulincer and P. R. Shaver, 203–218. American Psychological Association.

Heron-Delaney, M., S. Wirth, and O. Pascalis. 2011. "Infants' Knowledge of Their Own Species." *Philosophical Transactions of the Royal Society B: Biological Sciences* 366, no. 1571: 1753–1763.

Herrmann, E., J. Call, M. V. Hernández-Lloreda, B. Hare, and M. Tomasello. 2007. "Humans have Evolved Specialized Skills of Social Cognition: The Cultural Intelligence Hypothesis." *Science* 317, no. 5843: 1360–1366.

Hoyo, Á., M. R. Rueda, and R. Rodríguez-Bailón. 2019. "Children's Individual Differences in Executive Function and Theory of Mind in Relation to Prejudice Toward Social Minorities." *Frontiers in Psychology* 10: 2293.

Hui, M. M., and K. T. Kung. 2025. "Experimental Evidence of Peer Gender Nonconformity Triggering Dehumanization in Children: Developmental Trajectory, Form, and Link to Bullying." *Developmental Science* 28, no. 6: e70070.

Imuta, K., J. D. Henry, V. Slaughter, B. Selcuk, and T. Ruffman. 2016. "Theory of Mind and Prosocial Behavior in Childhood: A Meta-Analytic Review." *Developmental Psychology* 52, no. 8: 1192.

Jack, A. I., A. J. Dawson, and M. E. Norr. 2013. "Seeing Human: Distinct and Overlapping Neural Signatures Associated With Two Forms of Dehumanization." *Neuroimage* 79: 313–328.

Krupenye, C., F. Kano, S. Hirata, J. Call, and M. Tomasello. 2016. "Great Apes Anticipate That Other Individuals Will Act According to False Beliefs." *Science* 354, no. 6308: 110–114.

Kteily, N., G. Hodson, and E. Bruneau. 2016. "They See Us as Less Than Human: Metadehumanization Predicts Intergroup Conflict *via* Reciprocal Dehumanization." *Journal of Personality and Social Psychology* 110, no. 3: 343.

Kteily, N. S., and E. Bruneau. 2017. "Darker Demons of Our Nature: the Need to (re) Focus Attention on Blatant Forms of Dehumanization." *Current Directions in Psychological Science* 26, no. 6: 487–494.

Kteily, N. S., E. Bruneau, A. Waytz, and S. Cotterill. 2015. "The Ascent of Man: Theoretical and Empirical Evidence for Blatant Dehumanization." *Journal of Personality and Social Psychology* 109, no. 5: 901–931.

Kteily, N. S., and A. P. Landry. 2022. "Dehumanization: Trends, Insights, and Challenges." *Trends in Cognitive Sciences* 26, no. 3: 222–240.

Liu, D., H. M. Wellman, T. Tardif, and M. A. Sabbagh. 2008. "Theory of Mind Development in Chinese Children: A Meta-Analysis of False-Belief Understanding Across Cultures and Languages." *Developmental Psychology* 44, no. 2: 523.

Lolliot, S., K. Schmid, M. Hewstone, A. Al Ramiah, N. Tausch, and H. Swart. 2013. "Generalized Effects of Intergroup Contact: The Secondary Transfer Effect". In *Advances in intergroup contact*, 81–112. Psychology Press.

Loughnan, S., N. Haslam, and Y. Kashima. 2009. "Understanding the Relationship Between Attribute-Based and Metaphor-Based Dehumanization." *Group Processes & Intergroup Relations* 12, no. 6: 747–762.

MacLean, E. L., and B. Hare. 2012. "Bonobos and Chimpanzees Infer the Target of Another's Attention." *Animal Behaviour* 83, no. 2: 345–353.

McLoughlin, N., and H. Over. 2017. "Young Children Are More Likely to Spontaneously Attribute Mental States to Members of Their Own Group." *Psychological Science* 28, no. 10: 1503–1509.

Mcloughlin, N., S. P. Tipper, and H. Over. 2018. "Young Children Perceive Less Humanness in Outgroup Faces." *Developmental Science* 21, no. 2: e12539.

Meltzoff, A. N. 1995. "Understanding the Intentions of Others: Re-Enactment of Intended Acts by 18-month-old Children." *Developmental Psychology* 31, no. 5: 838.

Miller, S. A. 2009. "Children's Understanding of Second-Order Mental States." *Psychological Bulletin* 135, no. 5: 749.

Milligan, K., J. W. Astington, and L. A. Dack. 2007. "Language and Theory of Mind: Meta-Analysis of the Relation Between Language Ability and False-Belief Understanding." *Child development* 78, no. 2: 622–646.

Morera, M. D., M. N. Quiles, A. D. Correa, N. Delgado, and J. P. Leyens. 2018. "Perception of Mind and Dehumanization: Human, Animal, or Machine?" *International Journal of Psychology* 53, no. 4: 253–260.

Osterhaus, C., and S. L. Bosacki. 2022. "Looking for the Lighthouse: A Systematic Review of Advanced Theory-of-Mind Tests Beyond Preschool." *Developmental Review* 64: 101021.

Over, H. 2018. "The Influence of Group Membership on Young Children's Prosocial Behaviour." *Current Opinion in Psychology* 20: 17–20.

Over, H. 2021. "Seven Challenges for the Dehumanization Hypothesis." *Perspectives on Psychological Science* 16, no. 1: 3–13.

Pauen, S. 2000. "Early Differentiation Within the Animate Domain: Are Humans Something Special?" *Journal of Experimental Child Psychology* 75, no. 2: 134–151.

Rai, T. S., P. Valdesolo, and J. Graham. 2018. "Reply to Fincher et al.: Conceptual Specificity in Dehumanization Research Is a Feature, Not a Bug." *Proceedings of the National Academy of Sciences* 115, no. 15: E3331–E3332.

Rhodes, M. 2012. "Naïve Theories of Social Groups." *Child Development* 83, no. 6: 1900–1916.

Simon, J. C., and J. N. Gutsell. 2021. "Recognizing Humanity: Dehumanization Predicts Neural Mirroring and Empathic Accuracy in Face-to-Face Interactions." *Social Cognitive and Affective Neuroscience* 16, no. 5: 463–473.

Sullivan, K., D. Zaitchik, and H. Tager-Flusberg. 1994. "Preschoolers Can Attribute Second-Order Beliefs." *Developmental Psychology* 30, no. 3: 395.

Tajfel, H., M. G. Billig, R. P. Bundy, and C. Flament. 1971. "Social Categorization and Intergroup Behaviour." *European Journal of Social Psychology* 1, no. 2: 149–178.

Thomas, A. J., L. Thomsen, A. F. Lukowski, M. Abramyan, and B. W. Sarnecka. 2018. "Toddlers Prefer Those Who Win but Not When They Win by Force." *Nature Human Behaviour* 2, no. 9: 662–669.

Tomasello, M. 2018. "How Children Come to Understand False Beliefs: A Shared Intentionality Account." *Proceedings of the National Academy of Sciences* 115, no. 34: 8491–8498.

Tomasello, M. 2019. *Becoming Human: A Theory of Ontogeny*. Harvard University Press.

Tomasello, M., and H. Rakoczy. 2003. "What Makes Human Cognition Unique? From Individual to Shared to Collective Intentionality." *Mind & Language* 18, no. 2: 121–147.

van Noorden, T. H., G. J. Haselager, A. H. Cillessen, and W. M. Bukowski. 2014. "Dehumanization in Children: The Link With Moral Disengagement in Bullying and Victimization." *Aggressive Behavior* 40, no. 4: 320–328.

Wang, Z., R. T. Devine, K. K. Wong, and C. Hughes. 2016. "Theory of Mind and Executive Function During Middle Childhood Across Cultures." *Journal of Experimental Child Psychology* 149: 6–22.

Waytz, A., and N. Epley. 2012. "Social Connection Enables Dehumanization." *Journal of Experimental Social Psychology* 48, no. 1: 70–76.

Waytz, A., and J. Schroeder. 2014. "Overlooking Others: Dehumanization by Comission and Omission." *TPM-Testing, Psychometrics, Methodology in Applied Psychology* 21, no. 3: 251–266.

Wellman, H. M., D. Cross, and J. Watson. 2001. "Meta-Analysis of Theory-of-Mind Development: The Truth About False Belief." *Child Development* 72, no. 3: 655–684.

Wellman, H. M., and D. Liu. 2004. "Scaling of Theory-of-Mind Tasks." *Child Development* 75, no. 2: 523–541.

Wobber, V., E. Herrmann, B. Hare, R. Wrangham, and M. Tomasello. 2014. "Differences in the Early Cognitive Development of Children and Great Apes." *Developmental psychobiology* 56, no. 3: 547–573.

Woodward, A. L., J. A. Sommerville, and J. J. Guajardo. 2001. "How Infants Make Sense of Intentional Action." In *Intentions and Intentionality: Foundations of Social Cognition*, 149–169. The MIT Press.

Wright, S. C., A. Aron, and L. R. Tropp. 2014. "Including Others (and groups) in the Self: Self-Expansion and Intergroup Relations." In *The Social Self*, 343–363. Psychology Press.

Yu, H., X. Hu, Y. He, W. Li, and X. Mai. 2024. "Preschoolers' In-Group Bias Promotes Altruistic Sharing and Reduces Second-Party Punishment: The Role of Theory of Mind." *Journal of Experimental Child Psychology* 246: 106015.

Yu, J., L. Zhu, and A. M. Leslie. 2016. "Children's Sharing Behavior in Mini-Dictator Games: The Role of In-Group Favoritism and Theory of Mind." *Child Development* 87, no. 6: 1747–1757.

Zhan, Z. 2012. *A Study on Perceptual Reasoning and Verbal Ability Scale of 3-6 Years Old Children in Shanghai* [Master's Thesis, Shanghai Normal University]. https://kns.cnki.net/kcms2/article/abstract?v=iLvembebNjwUEUuRkRQjUqmCFhEWfNihKCLSAvmUjZPdZMuZq_SlrzflCD3DStUOJsEzqGfusB72ZIhEjlJK9S8_O7aFyN6lAtM5sg4WmicZpyYpYpFtKK0ErLGUCK8r46JCkR2SHI4oUwFxmzFZLVxwbIFg9MdxfdNNhibSSqEu17MGdXJ3d2tMMnGh4XTS&uniplatform=NZKPT&language=CHS.

Zhou, W. 2022. *Does our Perception of Animals Shape When we see all Humans as Being Created Equally?* [Doctoral diss., Duke University].

Zhou, W., A. Bowie, J. Tan, and B. Hare. 2024. "Humanizing Animals Does Not Reduce Blatant Dehumanization by Children or Adults." *Current Research in Ecological and Social Psychology* 6: 100194.

Zhou, W., and B. Hare. 2022. "The Early Expression of Blatant Dehumanization in Children and Its Association With Outgroup Negativity." *Human Nature* 33, no. 2: 196–214.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.

Supporting File 1: desc70165-sup-0001-SupMat.docx