Morally-Relevant Theory of Mind Mediates the Relationship between Group Membership and Moral Judgments

Jacquelyn Glidden, Alexander D'Esterre, and Melanie Killen

University of Maryland

Citation:
Glidden, J., D'Esterre, A., & Killen, M. (2021). Morally-relevant theory of mind mediates the relationship between group membership and moral judgments. *Cognitive Development.* https://doi.org/10.1016/j.cogdev.2020.100976

Correspondence concerning this article should be addressed to Jacquelyn Glidden, Department of Human Development and Quantitative Methodology, University of Maryland, College Park, 3942 Campus Drive, Suite 3304, College Park, MD 20742-1131. E-mail: jglidden@umd.edu.

The last author was supported by a grant from the National Science Foundation, BCS 1728918 and the National Institutes of Health, R01HD093698.

During everyday interactions, children are tasked with inferring others' mental states—their beliefs, desires, intentions, and motivations. The ability to accurately infer others' mental states, referred to as theory of mind (ToM), develops during the first decade of life and continues through adolescence (Hughes & Devine, 2015; Lagattuta, 2005; Wellman & Liu, 2004). Children experience considerable changes in their ToM skills across development (Wellman et al., 2001). Over the past two decades, researchers who study children's ToM skills have investigated the relations between ToM, social interactions, and social and moral knowledge. Carpendale and Lewis (2004), for example, proposed that children's developing understanding of the mind is not solely an individualistic process and that it occurs within triadic interactions involving children's experience of the social world, their communicative interactions with others, and moral knowledge. The current study seeks to bring together different lines of research to investigate how children's group membership and ingroup biases are related to their ToM (also more referred to more broadly as mental state understanding) and two types of person judgments: attributions of intentions and social exclusion.

Group Identity and Person Judgments

Contextual factors such as group identity (often studied as ingroup preference/outgroup dislike) have been shown to directly influence children's moral evaluations and behavior (Griffiths & Nesdale, 2006; Verkuyten, 2007). Children's tendency to favor their own social groups over groups to which they do not belong can affect their resource allocation and social exclusion/inclusion decisions (Abrams et al., 2003; Killen & Cooley, 2014; Mulvey, 2016).

It is particularly important to investigate how group identity can affect person judgments such as attributing intentions of others or decisions to socially exclude. Opportunities for these two types of judgments (intention attribution and exclusion) occur routinely throughout

childhood. Past research has shown that group membership can impact children's attribution of intentions. Specifically, McGlothlin and Killen (2006) investigated the racial biases of European American elementary school children in ethnically homogenous schools by showing children illustrations which were intentionally ambiguous, and which could be interpreted such that a character was either being helpful or committing a transgression. Importantly, there were two versions of each vignette, one in which the potential transgressor was a racial ingroup member (European American) and the potential victim was a racial outgroup member (African American), and a second version where the potential transgressor was a racial outgroup member (African American) and the potential victim was a racial ingroup member (European American). In line with the predictions of this study, the European American participants were more likely to assume helpful intentions when the potential transgressor was a racial ingroup member than when the potential transgressor was a racial outgroup member. This study suggests that group identity plays an important role in children's ability to attribute intentions of others.

Similarly, group identity also plays an important role in children's decisions to exclude peers, another type of person judgment. Work by Nesdale and colleagues (2005) examined the effects of levels of ingroup identification and outgroup threat on children's preferences for ethnic ingroups and outgroups (i.e., Pacific-Islanders vs. Anglo-Australian) in a sample of 6-9 year old Anglo-Australians. This work revealed that young children show negativity toward the outgroup, particularly when they strongly identify with their ingroup and experience threat from that outgroup. Importantly, very young children can also be subject to these intergroup biases, especially when it comes to gender. Work investigating the role of group identity in gender-based exclusion showed that children as young as 3.5 years of age are capable of identifying social exclusion based on gender, and they are able to use complex reasoning to reject such

exclusion (Killen et al., 2001). Ultimately, exclusion manifests when children are put in situations that encourage them to maintain positive ingroup identity and group functioning, including in competitive intergroup contexts.

While a large portion of the literature shows that group identity from social groups (e.g., race, gender) impacts person judgments, there is also evidence from the minimal and competitive groups literature showing the same pattern. Previous work using competitive school and sports teams replicate work with real social groups by showing that children's ingroup biases come into play even in minimal or temporary groups (McGuire et al., 2015, 2017). Interactions between group identity, mental state understanding, and moral cognition are most likely to occur in scenarios where all three elements are present and highly salient. Previous research has indicated that competition increases the importance of group identity and influences children's resource allocation behavior (McGuire et al., 2017). Misunderstandings and false accusations in competitive contexts are something that children are likely to be familiar with, making it an ideal context to use for separating out the roles of group identity, moral cognition, and mental state understanding. It has also been shown that children with more advanced mental state understanding were more likely to expect others to challenge group based cognition (Mulvey et al., 2016) and therefore there is reason to believe that mental state understanding could also serve as a buffer against the effects of ingroup favoritism or bias in these high-stakes competitive scenarios. Given that a competitive context seems to heighten concerns for group affiliation and can influence children's fairness decisions this context provided an interesting situation to introduce mental state information.

Person judgments, such as attributing intentions of others and deciding whom to exclude, can occur in intergroup contexts, making them subject to group identity processes, such as

ingroup bias or preference. It is particularly important to understand how group identity factors contribute to person judgments given the increasingly diverse environments of children and adults today. While there is evidence that group identity impacts person judgments and moral decision making, the mechanism for the relationship is unclear. Here, we hypothesize that mental state understanding, morally-relevant Theory of Mind (MoToM) specifically, serves as a mediator between group identity and children's person judgments. It is our belief that children's ability to take perspectives of outgroup members in morally-salient contexts can reduce the effect of ingroup biases on their person judgments, such as attributing intentions to outgroup members and deciding whether or not outgroup members should be excluded.

Morally-relevant Theory of Mind

Understanding others' intentions is a core aspect of both moral judgment and ToM (Lagattuta & Weller, 2014). Children's ToM abilities and moral judgments seem to have a bidirectional relationship, with their intentionality understanding being necessary for moral judgments, and children's moral judgments able to influence their interpretations of others' intentions (Leslie et al., 2006). Killen and colleagues (2011) investigated the intersection of mental state understanding and moral judgments by examining children's "morally-relevant" theory of mind (MoToM). The authors developed a false belief task that added salient social information to each step of the prototypic false belief scenario. In a prototypic false belief scenario, children witness a change in location of an object and must infer that someone who did not witness the change of location will look in an incorrect location. The MoToM measure developed by Killen and colleagues (2011) modified this task by embedding the change of location task in a socially salient context: a special cupcake in a lunch bag was transferred from a table to a trash can (while the owner was out of the room). The transfer was made by a socially

meaningful agent, that is, a classroom helper who did not know what was in the bag. In this task, assessments included these MoToM items (location change and false contents) as well as moral judgment items (evaluation of the act and assignment of blame). The findings revealed that MoToM skills served as better predictors of children's attributions of intentions of the accidental transgressor than did the prototypical ToM skills. This line of work has continued, with mounting evidence supporting the proposition that morally-relevant measures of children's mental state understanding provides an informative context-specific measure of children's ToM skills in situations involving moral intentions.

One study which has further supported and extended this literature investigated unintentional and intentional false statements regarding claims to resources (D'Esterre et al., 2019). In this study, MoToM was a better predictor than prototypic false belief ToM for children's (4-10 years old) differentiation of intentional and unintentional falsehoods. An intentional falsehood was a character making a claim to a necessary resource when they were fully aware that they didn't need the resource and an unintentional falsehood was making the same claim but not realizing that one already had the necessary resources. In both cases, making the false claim to the resource deprived another child who had a legitimate need, but the difference was in the knowledge state of the claimant. In this context, children's prototypical ToM competence predicted more favorable evaluations of the individual who made the unintentional false claim than the one who made an intentional false claim. Their MoToM ability, however, predicted children's responses to children's evaluations, attributions of intentions, and the assignment of punishment above and beyond age and prototypic ToM.

Taken together with the study by Killen and colleagues (2011), these studies provide support for the proposition that children's MoToM skills are better at predicting children's social

understanding when compared to scores on prototypical ToM measures. Overall, MoToM measures are better predictors than ToM skills in social contexts and provide a more ecologically valid measure by asking children to attribute mental states in socially and morally complex scenarios, especially when compared to traditional ToM measures, such as the Sally Ann task. Thus, to measure children's mental state understanding, the current study measured children's MoToM competence.

A growing body of research has shown that while ToM abilities improve with age, a child's age is not a perfect predictor of their ToM competency. Longitudinal work by Smetana and colleagues (2012), for example, shows that children experience varying rates of ToM acquisition and other recent work has shown that, when predicting moral judgments, ToM serves as a better predictor than age, and that MoToM is a better predictor than both age and prototypic ToM (D'Esterre et al., 2019).

In the current study, we were interested in how the relationship between group identity and children's person judgments might be mediated by MoToM skills specifically. Given the focus of this special issue on role of cognition in person judgments, the current study focuses on the underlying mechanism for person judgments, which was MoToM skills. Our wide age range (4-10 years old) allowed us to include data from children who have not yet mastered MoToM skills and those that are completely competent in the MoToM skills needed for these person judgments. This allowed us to measure the relationship more accurately across all skill levels and to reflect children's experiences more accurately.

A central goal of the current study was to examine the relations between mental state understanding, group membership (e.g., ingroup preference), and person judgments in the form of attributions of intentions and decisions to exclude a peer. We intentionally do not include age

in our analyses or models because we are focused on the underlying mechanism of change, which the authors believe is MoToM skills. It is proposed that the interaction between mental state understanding and group membership influence children's person judgments. We propose that mental state understanding is a critical skill in intergroup contexts and bears on the manifestation of ingroup biases within those contexts. Further, we believe that children's MoToM skills serve as a specific and ecologically valid measure of mental state understanding within the competitive intergroup contexts presented.

Mental State Understanding in Intergroup Contexts

Although there have been numerous studies on the relationship between children's mental state understanding and their moral judgments, there have been relatively few studies that investigate this relationship in *intergroup contexts.* Intergroup contexts are defined as situations where group membership is salient and attitudes about the ingroup or outgroup can result in biased judgments such as ingroup preference or outgroup distrust (Dunham et al., 2011; McLoughlin & Over, 2017; Rutland & Killen, 2015). Examining the relations between mental state understanding, morality, and intergroup attitudes raises new questions about how group membership influences one's recognition of others' intentions.

In one study by Mulvey, Bucheister, and McGrath (2016), researchers showed that increased ToM competence was an important predictor of children's evaluations of resource inequalities in an intergroup context. Children (3-6 years old) with higher ToM competence considered resource inequalities as unfair within an ingroup and within an outgroup, while children with lower ToM competence only considered resource inequalities as unfair when the inequality harmed an ingroup member (Mulvey et al., 2016). This study provides one example for how ToM contributes to an intergroup social decision regarding resource allocation.

Other intergroup social decisions are also influenced by varying levels of ToM competence. Specifically, second order ToM skills (i.e., thinking that others are thinking about what others are thinking) are related to children's inclusion and exclusion predictions. For example, Abrams and colleagues investigated the role of second order ToM in children's predictions of others' inclusion of deviant group members (Abrams et al., 2014). The authors found that second order ToM skills plays a critical role in 6-7 year old children's ability to infer and anticipate social inclusion using social cues in intergroup contexts. In this study, only children with higher second order ToM scores were able to discern differences between peers based on deviance from group norms and then use that evaluation to predict groups' differential preferences for a deviant ingroup or outgroup member. In this case, children's second order ToM skills directly predicted their ability to complexly analyze the situation according to peer status (e.g., deviance from group norm, and ingroup/outgroup status). This work provided compelling evidence that children's developing ToM skills play an active role in their ability to determine social inclusion and exclusion decisions, especially in intergroup contexts. The present study took this work one step further by hypothesizing a mediation model for the variables of interest. It was hypothesized that mental state understanding serves as a mediator between children's ingroup biases and their social exclusion decisions.

While second order ToM skills predicted children's social exclusion decisions, it has also been shown that other types of ToM measures can also predict children's person judgments. An additional type of ToM task is theory of social mind. In this task, children are required to set aside their own knowledge and attitudes towards a character to accurately understand how someone with less knowledge will react. For example, a child may know that a character stole toys from another character, and to pass the task the child would have to say that the second

character (who does not know their toys have been stolen) will act positively towards the thief. Researchers have investigated the role of theory of social mind in older children's (6-11 year old) predictions of group inclusion of deviant and normative ingroup and outgroup members (Abrams et al., 2009). Here, the authors found that children who performed better on the theory of social mind task used group memberships of others as a cue to judge how these others would feel towards deviant and normative members. Further, the authors included a measure of children's understanding of group norms and were able to demonstrate that theory of social mind predicted inclusion decisions more than understanding of group norms alone. It is not just children's understanding of group norms in intergroup contexts that contributes to their intergroup person judgments and predictions; ToM skills are contributing as well.

These three studies provide compelling evidence that both mental state understanding and group membership play a critical role in children's evaluations of fairness and social inclusion/exclusion decisions. Importantly, these studies used different types of ToM measures, but did not include MoToM. In the present study it was hypothesized that the same processes are at work and that MoToM skills are the best predictor. Specifically, we hypothesized that mental state understanding (operationalized as MoToM skills) serves as a mediator between children's ingroup biases and their person judgments (e.g., attribution of intentions and social exclusion) in intergroup contexts.

Group Membership and Context Affect Theory of Mind Competence

While mental state understanding is clearly important for children's intergroup interactions, it is also important to consider how intergroup situations can elicit changes to children's mental state understanding. Recent research shows that the context of children's intergroup experiences is related to their ability to use mental state understanding. When children

(3-7 year old) are placed in groups based on their gender and then find their team disadvantaged due to factors outside of their control (e.g., denied resources because their team are girls/boys), they are better at accurately identifying the mental states of others via false belief and belief emotion ToM skills (Rizzo & Killen, 2018). The researchers interpreted this finding as revealing an aspect of mental state understanding that is motivational in nature; children who were disadvantaged because of their gender were more likely to read the social context carefully, particularly considering the false beliefs and belief emotions held by others in the same situation. Children who were contextually disadvantaged performed better on prototypic ToM tasks than did children who were contextually advantaged. This work suggests that social context plays an important role in children's ToM abilities. In the current study, social contexts were varied in order to gain more information on the role of context in the proposed mediation models. Specifically, three different contexts were used, varying in morality and underlying intentions.

Other work shows that children's group membership impacts their mental state understanding. A recent study found that children's mental state understanding differed as a function of whether the target in question was an ingroup or outgroup member (Gönültaş et al., 2019). For this study, nine- to thirteen-year-old children were interviewed and presented with stories depicting characters in situations that involved misunderstandings, deception, and other contexts in which an understanding of beliefs and intentions would be critical. While participants did not differ in their general reasoning abilities or any other control variables, they did differ significantly in their mental state accuracy. Specifically, children were better at recognizing and correctly reasoning about the mental states of characters who shared a group membership with the participant than they were when presented with children of differing identities. In this study

there was strong evidence for the claim that children's understanding of mental states can be impacted by their group membership and viewing others as members of the ingroup or outgroup.

There is also evidence that group membership influences children's mental state attributions early in development and can influence the way young children reason about mental states. McLoughlin and Over (2017) found that 5- and 6- year old children spontaneously use more mental state words when talking about ingroup members compared to outgroup members. Further, 6-year-olds not only used more mental state words when talking about ingroup members, but they also used a greater diversity of mental state words when discussing ingroup members. This difference in children's attributions of mental states to outgroup members can be seen as an early form of ingroup bias. With the evidence from Gönültaş and colleagues (2019), there is a strong connection between children's group membership and their mental state understanding, such that ingroup biases can negatively affect children's abilities to infer others' mental states.

These findings reveal that children's group membership impacts their mental state understanding. Alongside the work showing that mental state understanding is critical for intergroup social decisions, it becomes clear that all three components (group membership, mental state understanding, and moral outcomes) must be considered together. The current study brought together several lines of research to investigate the roles of children's group membership and their MoToM skills in their person judgments: attributions of intentions to characters and children's own decisions to exclude characters from an intergroup context.

The Current Study

The goal of this study was to investigate whether children's MoToM competence influences the extent to which children's ingroup biases impact their ability to attribute intentions

to others and decide whether to exclude someone from a team, across multiple complex contexts. Children were asked to evaluate three different vignettes in which target children created one of three advantage types for their team: 1) unintentional unfair advantage, 2) intentional fair advantage, or 3) intentional unfair advantage. These three situations varied in moral and intentional outcomes. In the *unintentional unfair* advantage condition, the protagonist did not know they violated a rule and gained an advantage. In the *intentional fair* advantage condition, the protagonist created an advantage simply by adhering to the rules of the contest when the opposing team failed to do their best. In the *intentional unfair* advantage condition, the protagonist intentionally violated a rule, creating a straightforward transgression. These three contexts were investigated to determine how the relationship between group membership, MoToM competence, and children's person judgments differed across contexts that varied in protagonist intentions and moral outcomes. This design provided a basis for teasing apart the influence of positive and negative intentions regarding evaluations of rule transgressions.

While both MoToM competence and Attribution of Intentions (AoI) both reference intentions, they are two distinct but related processes. AoI is an evaluation assessment and not a mental state understanding question, requiring participants to address the acceptability of someone's intentions. In the context of this study, for example, participants can fail to understand the intentions of a character and still give a high AoI evaluation (e.g., "they thought it was really good because they wanted to help their team win") or participants can give a negative AoI even if intentions are not understood (e.g., "They think they're doing something pretty bad because they know they didn't double check with their team before they fed the pumpkins"). Within the context of this study, an understanding of intentions will influence participants' evaluations of character intentions, but the two are distinct processes (MoToM: understanding

vs. AoI: evaluation). Thus, in this study, MoToM was used as a predictor and mediator, while AoI was a person judgment (acceptability of the characters' intentions).

To test the goals of the study, participants were inducted into one of two teams, the red or the blue team, using Nesdale's (2004) group membership procedure which involved making group membership salient (see Methods). The key outcome measures were participants' attributions of intentions of the advantage creator, and participants' decision to exclude the advantage creator from the team. These two distinct person judgments allow the authors to determine if MoToM competence serves as a mediator across various forms of person judgments. While AoI is evaluative of mental states, decisions to exclude focus on character traits and consequences (i.e. worthiness of exclusion, deserving of a social consequence). Participants' MoToM competence was measured within each story (e.g., by asking whether the participant thought that the target child had access to necessary information).

It was hypothesized that MoToM would mediate the relationship between group membership and the outcome measures, and further, that this relationship would vary across the three advantage contexts (see Figure 1 for proposed mediation model). Previous research has demonstrated that ToM and MoToM competence is necessary for children's person judgments (D'Esterre et al., 2019; Killen et al., 2011; Mulvey et al., 2016), but no studies have investigated whether group membership is related to MoToM skills in an intergroup context. Thus, the first step was to determine whether children's group membership would predict their MoToM competence. After identifying a meaningful relationship between group membership and MoToM competence, mediation analyses were conducted. The hypotheses were as follows:

H1) *Group membership would predict children's MoToM competence*. Based on recent research suggesting that children may be more sensitive to the mental states of ingroup members

(e.g., Gönültaş et al., 2019; McLoughlin & Over, 2017) we predicted that individuals would be more attentive to the mental states of a target when the target was identified as an ingroup member than when the target was identified as an outgroup member.

H2) *Group membership would predict children's attributions of a character's intentions, with ingroup members being more positive in their attributions, and that this relationship would be mediated by a child's MoToM performance*. This prediction of a mediated relationship was based on research showing ingroup preferences in children's intention attributions (McGlothlin & Killen, 2010) and connections between children's ToM performance and their moral judgments (e.g., Mulvey et al., 2016). It was also expected this relationship would vary based on the complexity of a character's behavior such that MoToM would mediate in contexts where intentions play an important role in framing outcomes, such as with an unintentional unfair advantage (H2a) and intentional fair advantage (H2b), but that MoToM would not mediate this relationship with an intentional unfair advantage (prototypic moral transgression) (H2c).

H3) *Group membership would also predict children's decisions to exclude characters from the competition, with ingroup members being less likely to exclude the advantage creator, and that this relationship would be mediated by a child's MoToM performance.* Previous work has demonstrated that children's advanced ToM skills are related to their social inclusion and exclusion decisions (Abrams et al., 2009, 2014), and that group membership impacts mental state understanding (McLoughlin & Over, 2017) and social exclusion decisions. Thus, it was also expected that MoToM competence would mediate this relationship in contexts containing an unintentional unfair advantage (H3a) and intentional fair advantage (H3b), but that this mediation relationship would not be present in the intentional unfair advantage, a straightforward moral transgression (H3c).

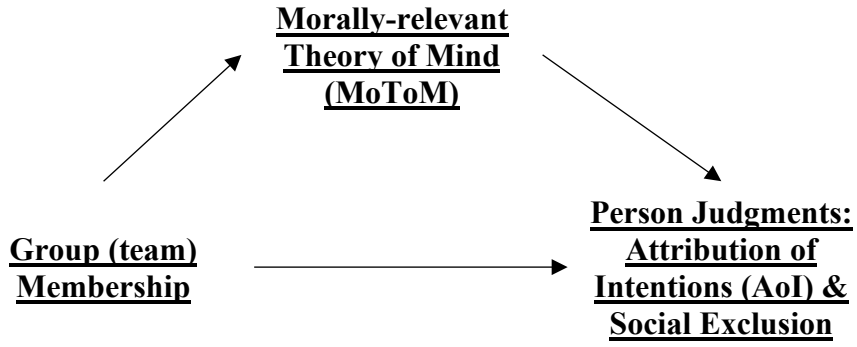

Figure 1. Proposed mediation model: Morally-relevant Theory of Mind (MoToM) mediates the relationship between group membership and person judgments.

Method

Participants

Participants included 120 children between 4 and 10 years of age (M_{Age} = 6.87 years, SD_{Age} = 1.81; 53% female) recruited from preschools and summer camps. Sample size was determined using a priori power analyses using G*Power (Faul et al., 2009), which revealed that in order to detect small to medium effects, a minimum of approximately 100 participants would be necessary to test our hypotheses. Participants were ethnically diverse (67% European American, 18% African American, 11% Asian American, and 4% Hispanic) and were recruited from preschools and summer camp serving lower-middle to upper-middle income families in the Mid-Atlantic region of the United States.

Design

In order to investigate the relationship between group membership and advantage type a 2 (group membership: ingroup, outgroup) x 3 (advantage type: unintentional unfair, intentional fair, or intentional unfair) mixed-factorial design was utilized with the group affiliation as a between-subject manipulation and advantage type as a within-subject manipulation. All analyses controlled for age.

Procedure

This project was approved by the Institutional Review Board at the University of Maryland entitled Children's Evaluations of Unintentional and Intentional Rule Violations in Intergroup Contexts [#1185366-2]. All participants received written parental consent to participate and gave verbal assent prior to study administration. Trained research assistants individually administered the task to all participants. Interviews were conducted in a quiet space in participants' schools and lasted approximately 15-20 minutes. The research assistants read the children stories from a script which was presented using a brightly illustrated PowerPoint presentation on a laptop computer. Researchers used a printed protocol to record children's Likert-type response, and all sessions were audiotaped. Participants heard three vignettes: one about an unintentional unfair advantage, one about an intentional fair advantage, and a third about an intentional unfair advantage. Participants heard stories in which the blue team member was the advantage creator. For participants on the blue team this was an ingroup member, and for participants on the red team, this was an outgroup member.

Participants were introduced to a 6-point Likert scale and were trained on its use. Once children demonstrated their comprehension of the scale and were able to reliably and comfortably use the midpoints and both endpoints, the researcher began the first vignette. All participants were presented with the stories in the same order (Unintentional Unfair, Intentional Unfair, Intentional Fair). Between each story participants discussed hobbies with the interviewer (e.g., favorite animal, favorite pastime, favorite movie). After this short break, participants were told there was going to be a new contest and introduced to their new teammates.

Group Membership. Before hearing about the advantage contexts, children were randomly assigned to either the ingroup or outgroup conditions. This team assignment process

served as the condition manipulation in which children assigned to the ingroup condition (blue team) heard advantage contexts in which the advantage creator was an ingroup member who helped their team, while those assigned to the outgroup condition (red team) were disadvantaged by an outgroup advantage creator in each context. In order to make team membership salient (Nesdale, 2004), children were asked to pick a star or lightning bolt icon for their team and to hold a small laminated picture of their chosen icon during the administration of the task. Further, children were also asked to select a hypothetical reward for their team if their team won the contest, taking the form of an ice cream or pizza party. All participants identified with their team and thought it would be better if their team won than if the other team won (*ps* < .001).

All children were presented with images of the characters on their own team and the other team on the PowerPoint display. On the participant's own team, a gender-matched silhouette character entitled "you" represented the participant. Each participant saw their silhouette standing with other characters wearing shirts that corresponded to their team, and all characters were portrayed as approximately the participant's own age and represented an ethnically varied team composition (see Figure 2). In between each story, children were shown a "filler task" which included separate power point slides with pictures of fun activities or images and asked how much they liked X (e.g., "How much do you like painting pictures? With a cartoon image of an easel and paints).

At the beginning of each condition, children were informed that the red and blue team were competing in a pumpkin growing contest and that the winner would get a cool prize. They were told that the main rule for this contest was that each team could only give their pumpkins *one cup of plant food* each day.

Unintentional Unfair Advantage. Previous research into children's ability to coordinate intentions and outcomes has shown that when intentions and outcomes are in opposition to one another (i.e. good intentions but a negative outcome) children in this age range often struggle to coordinate this competing information (D'Esterre et al., 2019). Therefore, the first advantage context we sought to create involved an unintentional transgression which resulted in an unfair outcome.

In the story about the unintentional unfair transgressor, children were introduced to a blue team character named Sam and the one cup of plant food per day rule. Children were told that it was Sam's turn to feed the pumpkins, but he could not find the plant food, so he left to look for it. While Sam was away his teammates found some plant food, both teams fed their pumpkins, and then everyone left. After everyone had left, Sam returned with the plant food that he found, and he proceeded to feed his team's pumpkins, resulting in a context where Sam intended to follow the rules but unknowingly created an unfair advantage. Children were then informed that the blue team grew the biggest pumpkin and won the contest. At this point, children were given a memory check question: "Is Sam on your team or is Sam on the other team?" Children who responded incorrectly had the manipulation repeated to them two or fewer times, and all children successfully passed this memory check.

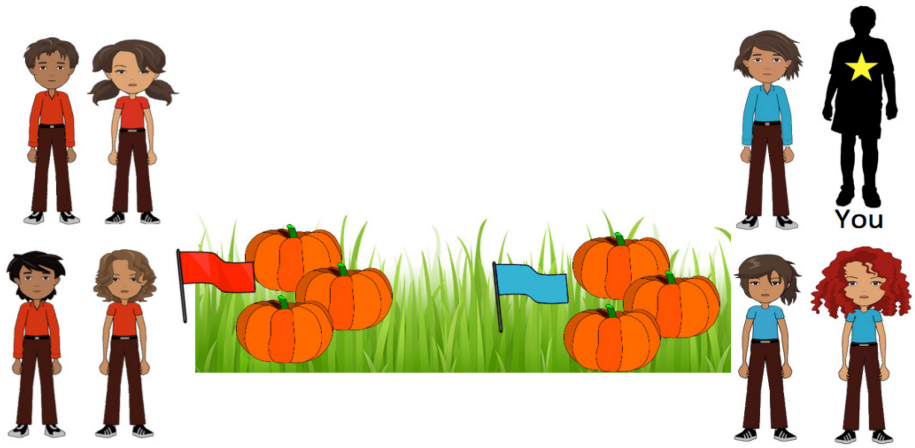

Figure 2. Power point slide depicting the team set-up for the competition when a male participant was assigned to the blue team (the silhouette was on the left side when the participant was assigned to the red team).

Intentional Fair Advantage. Similarly to the unintentional unfair advantage, the intentional fair advantage context was designed to create a context in which a character intended to follow the rules, but ultimately created an asymmetry in outcomes between the two competing teams. However, in this context no rules were violated and the advantage was created fairly through dedicated effort of a teammate. Previous developmental research has shown that children in this age range recognize the difference between fair and unfair advantages but that their evaluation of these advantages are influenced by whether it benefits them or another (Rizzo & Killen, 2018). This context is particularly interesting when contrasted with the unintentional unfair advantage context as both involve an intent to follow the rules, followed by an advantage being created for one team, but differ in whether the advantage was created fairly and intentionally.

In the intentional fair advantage story, children were introduced to Casey, a member of the blue team. Participants were reminded of the "one cup of plant food" rule and told that it was

Casey's turn to feed the pumpkins. Participants were informed that it was a very nice day outside and that the other kids decided to go to the park instead of feeding their pumpkins. Casey decided to stay and feed her team's pumpkins instead. Children were told that the blue team grew the biggest pumpkin and won the contest. Once again, a memory check was administered: "Is Casey on your team or is Casey on the other team?" Children who answered correctly proceeded to the next questions, whereas children who answered incorrectly had the manipulation repeated to them two or fewer times and all children successfully passed this memory check.

Intentional Unfair Advantage. Finally, we wanted to determine if children's responses to a prototypic moral transgression would show a different pattern from the other tested contexts. To this end the third and final context involved a character that knowingly broke a rule to create an advantage for their team, resulting in a cognitively simpler context in which children are not made to coordinate positive intentions (an effort to follow the rules) with a potentially negative outcome (one team advantaged over another). This context is interesting to compare to the previous two contexts because an advantage for one team is created, but here the character has negative intentions and their actions result in a negative outcome (an unfair advantage).

In the story about the intentional unfair advantage, children were introduced to Taylor, a blue team member, and reminded of the "one cup of plant food" rule. Participants were told that it was Taylor's turn to feed the pumpkins, but she could not find the plant food. After she looked around, she found the plant food, and both Taylor and the red team fed their pumpkins. Then, after everyone left, Taylor came back with another cup of plant food fed her team's pumpkins again. Previous research examining children's perceptions of intentions suggests that young children often mistakenly attribute forgetfulness as a rationale for intentional transgressions

(D'Esterre et al., 2019), and therefore it was stated that Taylor remembered that she fed the pumpkins earlier. Children were told that the blue team grew the biggest pumpkin and won the contest. At this point a memory check was asked: "Is Taylor on your team or is Taylor on the other team?" Children who answered correctly proceeded to the next questions, whereas children who answered incorrectly had the manipulation repeated to them two or fewer times, and once again all children successfully passed this memory check.

Measures

After finishing each story, in a fixed order, children were asked a *MoToM* question regarding the beliefs of the advantage creating character, an *attribution of intentions (AoI)* question regarding the motivation of the advantage creator, and a *social exclusion* question to decide whether or not the participant would suggest excluding the advantage creator from the contest. These two outcome measures (AoI and exclusion) were selected as they are two distinct person judgments which allowed the authors to determine if MoToM competence served as a mediator across various forms of person judgments. The two outcome measures provided insight into children's evaluations of intentions as well as their assessment of character traits (i.e., worthiness of social exclusion) due to the character's decisions. This resulted in six outcome measures and one mediation measure, and descriptive analyses revealed that participants utilized the entire range of these measures (Table 1).

Morally-relevant ToM (MoToM). At the end of each story participants were assessed on their perception of the advantage creator's knowledge state. To measure participants' MoToM competence, children's responses to first-order beliefs about the characters in the stories were used to create the MoToM assessment within each story. Specifically, participants were asked, "Does Sam/Casey/Taylor think that the pumpkins were already fed today?" Participants who

correctly indicated that 1) Sam did not think that the pumpkins were already fed 2) Casey did not think that the pumpkins were already fed and 3) Taylor *did* think that the pumpkins were already fed were indicated to have passed the respective question and their scores were reflected as a 1 if they passed and a 0 if they did not. Scores were then summed across all three scenarios to form a MoToM composite score.

Attribution of Intentions (AoI). The AoI assessment was administered by the researcher asking the participant "Does [Sam/Casey/Taylor] think he/she was doing something OK or not OK when he/she fed the pumpkins? How OK/not OK? Participants responded on a 6-point Likert scale ranging from 1 (*really not OK)* to 6 (*really OK).*

Social Exclusion. After participants were asked to attribute intentions to the advantage creator they were asked whether or not they would support excluding this character from future contests. The researcher asked the participant "Do you think [Sam/Casey/Taylor] should be off the team?" Participants responded using a binary Yes/No response.

Table 1. Descriptive Statistics for Mediator and Outcome Variables.

Measure	Context	Range	Mean	SD
Morally-relevant Theory of Mind Composite	Story Independent	0-3	2.39	0.74
Attribution of Intentions	Unintentional Unfair	1-6	4.33	1.72
	Intentional Fair	1-6	4.87	1.62
	Intentional Unfair	1-6	3.97	1.62
Social Exclusion	Unintentional Unfair	0-1	0.34	0.48
	Intentional Fair	0-1	0.17	0.37
	Intentional Unfair	0-1	0.42	0.50

Note. Within each context participants received the same morally-relevant Theory of Mind question, and their responses were summed to make a composite score. Total $N = 120$.

Data analytic plan

In order to explore the possibility of MoToM mediating the relationship between group membership and children's assessment of the advantage creators a series of linear regressions were conducted, with planned post-hoc t-tests in order to further illuminate findings of each regression. For the final mediation analysis, the "mediate" function from the "mediation" package in R version 3.6.1 was utilized, which provided us with the Average Causal Mediation Effects (ACME) for each of our models. This allowed us to state definitively whether MoToM performance significantly mediated the relationship between group membership and the two outcome variables (AoI and Exclusion) for each of our three advantage contexts. In each case group membership was entered as the treatment variable, MoToM performance was entered as the mediator, and 500 simulations with bootstrapping were utilized in order to ensure convergence.

Results

Group Membership Predicts MoToM Performance

In order to test our first hypothesis, that the assigned group membership of the child would significantly influence their MoToM performance (H1), we utilized a linear regression with team identity predicting MoToM scores. In support of our hypothesis, a significant effect of group membership was found ($F(1,118) = 5.50$, $p = .021$, $\beta = 0.31$) and planned post-hoc comparisons revealed that participants who shared a group membership with the advantage performed significantly better on the MoToM assessment ($M = 2.55$, $SE = 0.09$) than participants who did not share a group membership with the advantage creator ($M = 2.24$, $SE = 0.10$), $t(118) = 2.345$, $p = .021$ (Figure 3). Critically, this finding shows that children's MoToM competence

was impacted by their own group membership; if children shared team membership with the advantage creator then they were more likely to correctly answer the three MoToM questions.

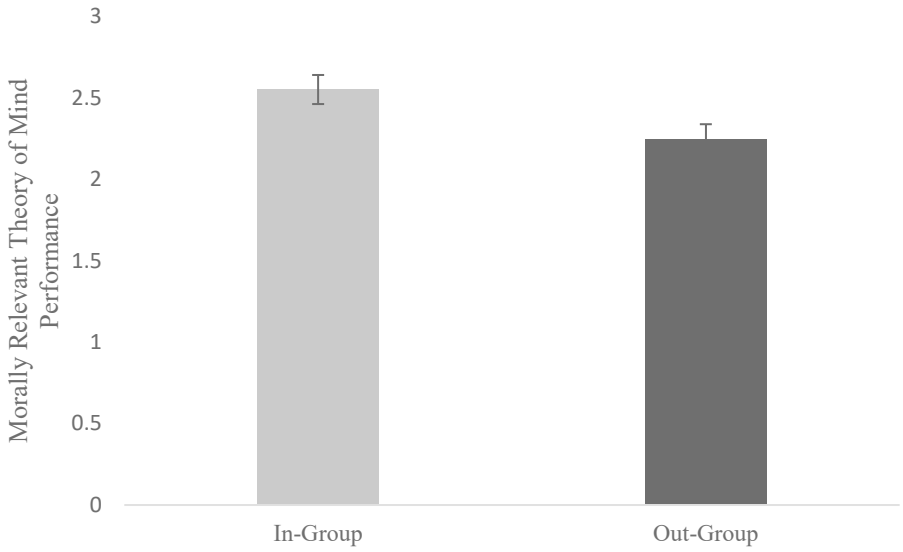

Figure 3. Participants responding about the beliefs of their ingroup team member performed better on the MoToM measure than did participants evaluating the outgroup team member.

Attributions of Intentions

Because it was predicted that MoToM would mediate the relationship between group membership and AoI for some contexts (H2a) (H2b) and not others (H2c), the results section for the second hypothesis has been divided into three smaller subsections which provide an in-depth focus for each condition. For each condition (unintentional unfair, intentional fair, and unintentional unfair) we first conducted a series of linear regressions to confirm that a significant relationship existed between group membership and AoI and between MoToM and AoI, before running the mediation analysis.

Unintentional Unfair Advantage. In the unintentional unfair advantage condition, it was hypothesized that MoToM performance would mediate the relationship between group

membership and AoI. To test this, a linear regression was conducted to determine the relationship between group membership and children's AoI. A significant relationship was found between group membership and AoI ($F(1,118) = 9.27$, $p = .003$, $\beta = 0.92$). Planned post-hoc comparisons revealed that participants who shared a group membership with the advantage creator attributed significantly more positive intentions ($M = 4.81$, $SE = 0.19$) than did participants who did not share a group membership with the advantage creator ($M = 3.89$, $SE = 0.24$), $t(118) = 3.045$, $p = .003$). Next a second linear regression was conducted to investigate the relationship between MoToM performance and AoI, and once again a significant relationship was found ($F(1,118) = 18.44$, $p < .001$, $\beta = 0.86$). The beta coefficients supported the prediction that participants who scored higher on the MoToM assessment would attribute significantly better intentions to the advantage creator. Finally, once each of the three relations were confirmed to be significant, the mediation analysis could be conducted to test the ability of MoToM to mediate the relationship between group membership and AoI.

Supporting the hypothesis that MoToM competence would mediate the relationship between group membership and AoI (H2a), a significant ACME was found for AoI (ACME = 0.2347 [0.0209, 0.52], $p = .02$) in the unintentional unfair advantage context (Figure 4A). Partial mediation was observed, with 25.42% of the direct effect mediated for the AoI measure, supporting the first mediation hypothesis (H2a), and showing that MoToM skills mediated the relationship between group membership and children's AoI in the unintentional unfair advantage context. While participants who shared a group membership with the advantage creator tended to attribute more positive intentions to the advantage creator than participants who did not share a group membership, this relationship was partially mediated by their MoToM performance, and the effect of group membership on AoI was significantly lower once this factor was incorporated

into the model. Thus, when participants heard about an unintentional unfair advantaged, MoToM skills mediated the relationship between their group membership (shared or not shared with the advantage creator) and their AoI for the advantage creator.

Intentional Fair Advantage. Next, the hypothesis that MoToM would mediate the relationship between group membership and AoI in the intentional fair advantage was examined (H2b). First, the relationship between group membership and AoI was tested with a linear regression and found to be significant ($F(1,118) = 4.34$, $p = .039$, $\beta = 0.61$), and planned post-hoc comparisons revealed that participants who shared a group membership with the advantage creator attributed significantly more positive intentions ($M = 5.19$, $SE = 0.18$) than participants who did not share a group membership with the advantage creator ($M = 4.58$, $SD = 0.23$), $t(118) = 2.084$, $p = .039$. Next, the relationship between MoToM performance and AoI was tested with a linear regression and a significant effect of MoToM performance on AoI was found ($F(1,118) = 9.53$, $p = .003$, $\beta = 0.60$) for the intentional fair context. The beta coefficient of this regression showed that participants who scored higher on the MoToM assessment attributed significantly more positive intentions to the advantage creator.

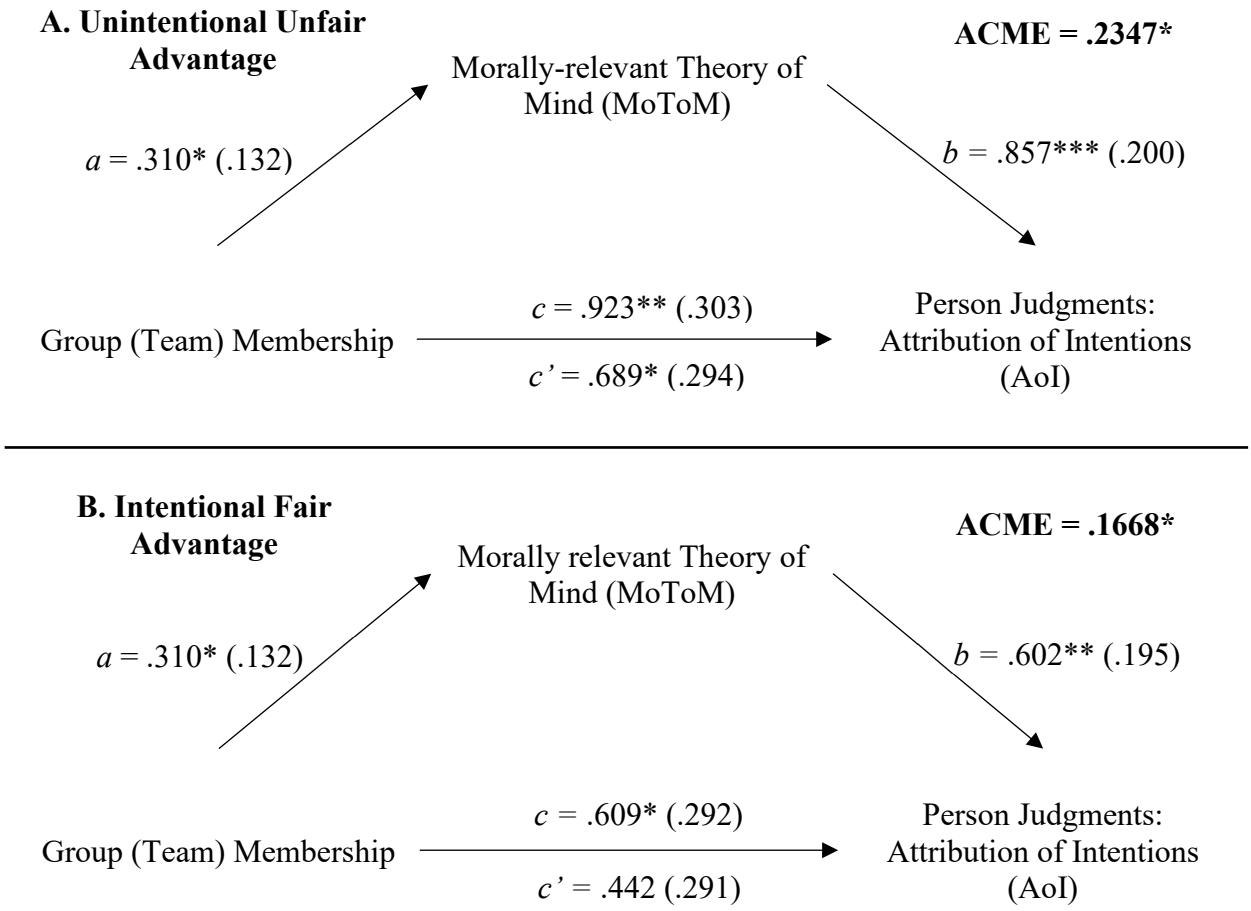

Figure 4. Mediation models showing that Morally-relevant Theory of Mind mediates the relationship between group membership and attribution of intentions in the A) Unintentional Unfair Advantage and B) Intentional Fair Advantage. * Significant at $p < .05$, ** significant at $p < .01$, and *** significant at $p < .001$.

Then the full mediation model was tested. Supporting the hypothesis of mediation (H2b), a significant ACME was found for the AoI (ACME = 0.1668 [0.0093, 0.42], $p = .032$, Figure 4B). Partial mediation was observed, with 27.38% of the direct effect mediated for the AoI measure. Thus, the second mediation hypothesis (H2b) was fully supported: MoToM

competence mediated the relationship between group membership and AoI in the intentional fair context. The same results were found as those in the unintentional unfair advantage: participants who shared a group membership with the advantage creator tended to attribute more positive intentions to the advantage creator than participants who did not share a group membership, this relationship was partially mediated by their MoToM performance, and the effect of group membership on AoI was significantly lower once MoToM was incorporated into the model. These results suggest that, in an intentional fair advantage context, MoToM skills meaningfully mediated the relationship between a participant's group membership and their AoI for the fair advantage creator.

Intentional Unfair Advantage. The ability of MoToM competence to mediate the relationship between group membership and AoI for the intentional unfair advantage (H2c) was investigated last. To test the relation between group membership and AoI, a linear regression was conducted, but a significant relationship was not found ($F(1,118) = 0.44$, $p = .507$, $\beta = 0.20$). Participants on the red and blue teams did not differ in their evaluations of the intentions of the intentional advantage creator. Likewise, when testing the relationship between MoToM performance and AoI, no significant effect was found ($F(1,118) = 0.181$, $p = .672$, $\beta = 0.09$). Participants' MoToM performance was not related to their AoI of an intentional unfair advantage creator. These results are in line with other research on children's moral judgments which suggest that even strong indicators of group membership do not change children's judgments about straightforward moral transgressions (e.g. D'Esterre et al., 2019). Because these relations were not significant, a mediation model did not hold for the intentional unfair advantage context.

Attributions of Intentions Summary. Overall, the hypotheses regarding MoToM mediating children's AoI were confirmed. Significant relations were found between group

membership and AoI for the unintentional unfair advantage and intentional fair advantage contexts, and participant's MoToM scores significantly mediated those relations (H2a and H2b). Also in line with the predictions, this relationship did not hold for the intentional unfair advantage (H2c), which further supports the idea that children's MoToM scores capture the ability of children to accurately infer intentions in complex moral scenarios, but that this ability does not provide any comprehension advantage in a simpler and more straightforward scenario.

Social Exclusion

MoToM competence mediated the relationship between group membership and the first person judgment, children's attributions of intentions. While previous research has shown differences and distinctions between MoToM competence and children's attributions of intentions (D'Esterre et al., 2019), it was not surprising that the two were related and that MoToM skills were directly involved in children's ability to determine intentions. Thus, tests were conducted to determine whether MoToM competence functioned as a mediator for a different person judgment, when the outcome variable was less directly related to intentions, such as children's decisions to socially exclude a peer. Previous work has shown that mental state understanding plays a critical role in children's understanding of social exclusion and their decisions to exclude peers (Abrams et al., 2009, 2014). Further, previous work also shows that children's group membership can impact exclusion decisions (Nesdale, Durkin, et al., 2005). Thus, we hypothesized that the complex relationship between these variables might be explained through a mediation model where MoToM skills mediated the relationship between children's group membership and their decisions to exclude.

It was predicted that MoToM performance would be a significant mediator between group membership and social exclusion for the unintentional unfair (H3a) and intentional fair

(H3b) contexts, but MoToM would not be a meaningful mediator for the intentional unfair context (H3c).

Unintentional Unfair Advantage. The first hypothesis tested was that MoToM performance would mediate the relationship between group membership and social exclusion in the unintentional unfair advantage (H3a). To test the relationship between group membership and exclusion, a linear regression was conducted and a significant effect was found ($F(1,118) = 43.96$, $p < .001$, $\beta = 0.49$). Planned post-hoc comparisons revealed that participants who shared a group membership with the advantage creator were significantly less likely to exclude ($M = 0.09$, $SE = 0.04$) than participants who did not share a group membership with the advantage creator ($M = 0.58$, $SE = 0.06$), $t(118) = 6.630$, $p < .001$. Following this, the relationship between MoToM performance and social exclusion in the unintentional unfair context was tested through another linear regression, and a significant effect was found for exclusion ($F(1,118) = 12.79$, $p = .001$, $\beta = -0.202$). The beta coefficient showed that participants who scored higher on the MoToM assessment were significantly less likely to condone exclusion.

The full mediation model was tested next. Supporting the hypothesis of mediation, a significant ACME waws found for the exclusion measure (ACME = -0.0425 [-0.1111, 0.00], $p = .036$, Figure 5A). Partial mediation was observed, with 8.60% of the direct effect mediated for the exclusion measure. Thus, the mediation hypothesis (H3a) was fully supported. MoToM skills mediated the relationship between group membership and children's social exclusion evaluations in the unintentional unfair context. While participants who did not share a group membership with the advantage creator were more likely to exclude the advantage creator than participants who did share a group membership, this relationship was partially mediated by their MoToM performance, and the effect of group membership on exclusion decisions was significantly lower

once this factor was incorporated into the model. Thus, when participants heard about an unintentional unfair advantaged, MoToM skills mediated the relationship between their group membership (shared or not shared with the advantage creator) and their exclusion decisions for the advantage creator.

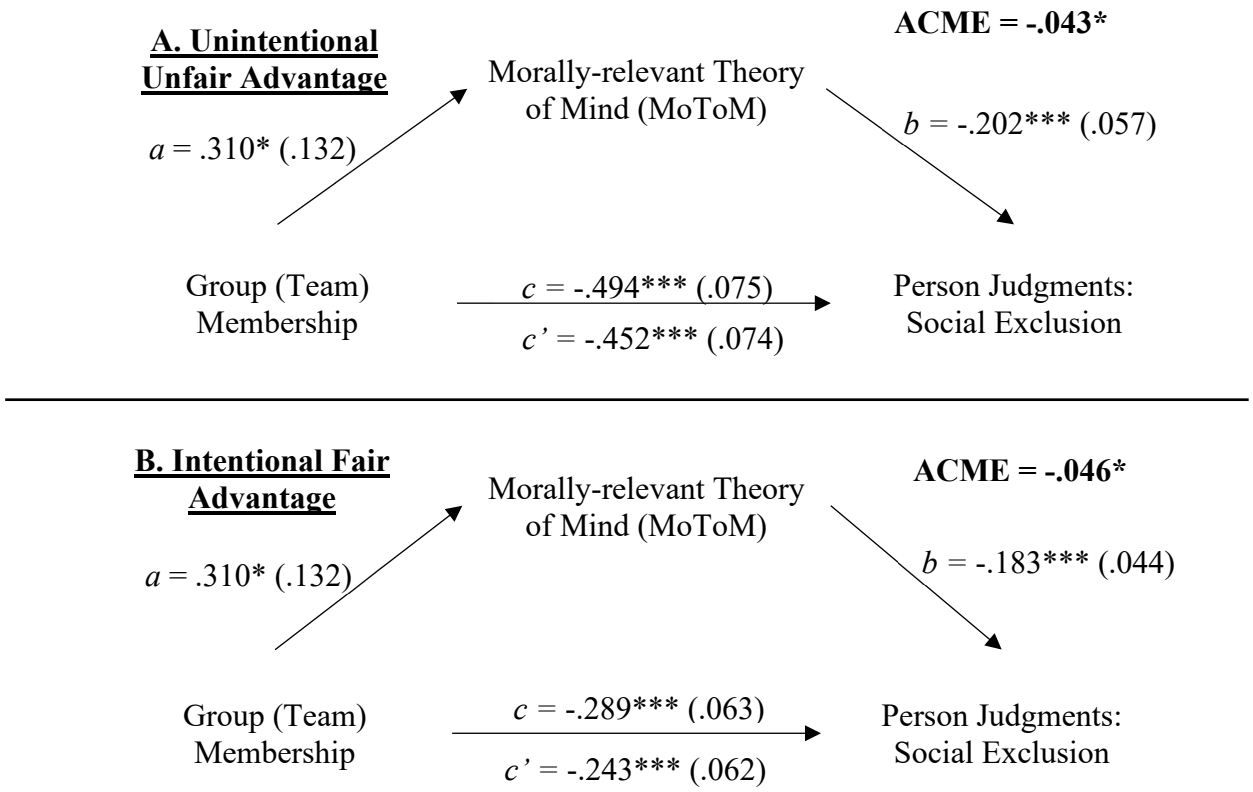

Figure 5. Mediation models showing that Morally-relevant Theory of Mind mediates the relationship between group membership and social exclusion in the A) Unintentional Unfair Advantage and B) Intentional Fair Advantage. * Significant at $p < .05$, ** significant at $p < .01$, and *** significant at $p < .001$.

Intentional Fair Advantage. Next, the ability of MoToM performance to mediate the relationship between group membership and exclusion decisions in the intentional fair advantage

(H3b) was examined. First, the relationship between group membership and exclusion was tested using a linear regression, which confirmed a significant effect ($F(1,118) = 20.89$, $p < .001$, $\beta = -0.29$) and planned post-hoc comparisons revealed that participants who shared a group membership with the advantage creator were significantly less likely to exclude ($M = 0.02$, $SE = 0.02$) than participants who did not share a group membership with the advantage creator ($M = 0.31$, $SE = 0.06$), $t(118) = 4.570$, $p < .001$. Next, the relationship between MoToM performance and exclusion for the intentional fair advantage context was tested and once again a significant effect was found ($F(1,118) = 17.64$, $p < .001$, $\beta = -0.18$). The beta coefficients supported the prediction that participants with higher MoToM competence were significantly less likely to condone exclusion.

The full mediation model was tested next. Supporting the hypothesis of mediation, a significant ACME was found for exclusion (ACME = -0.0459 [-0.1047, -0.01], $p = .036$, Figure 5B). Partial mediation was observed, with 15.88% of the direct effect mediated for the exclusion measure. Thus, the mediation hypothesis (H3b) was fully supported. MoToM skills mediated the relationship between group membership and children's exclusion evaluations in the intentional, fair context. While participants who did not share a group membership with the advantage creator were more likely to exclude the advantage creator than participants who did share a group membership, this relationship was partially mediated by their MoToM performance, and the effect of group membership on exclusion decisions was significantly lower once this factor was incorporated into the model. Thus, when participants heard about an intentional fair advantage, MoToM skills mediated the relationship between their group membership (shared or not shared with the advantage creator) and their exclusion decisions for the advantage creator.

Intentional Unfair Advantage. Next, the ability of MoToM to serve as a mediator in the relationship between group membership and participant's exclusion decision in the intentional unfair advantage (H3c) was examined. Unlike the analyses for children's AoI, when conducting a linear regression to investigate the relationship between children's group membership and their exclusion decision, a significant effect was found ($F(1,118) = 17.49$, $p < .001$, $\beta = -0.355$). Post-hoc comparisons revealed that participants who shared a group membership with the advantage creator were significantly less likely to exclude ($M = 0.24$, $SE = 0.06$) than participants who did not share a group membership with the advantage creator ($M = 0.60$, $SE = 0.06$), $t(118) = 4.182$, $p < .001$). The relationship between MoToM performance and exclusion was tested using a linear regression, but here no significant effect of MoToM performance was found ($F(1,118) = 0.57$, $p = .451$, $\beta = -0.05$). Because there was no significant relationship between MoToM performance and children's exclusion decisions, no significant mediation could be tested and MoToM could not be a significant mediator for the intentional unfair advantage context (H3c).

Social Exclusion Summary. Overall, a series of mediation hypotheses for social exclusion were confirmed. Significant relations were found between group membership and exclusion for the unintentional unfair advantage and intentional fair advantage contexts, and participant's MoToM scores significantly mediated those relations (H3a and H3b). Also in line with predictions, this relationship did not hold for the intentional unfair advantage (H3c), which further supports the idea that children's MoToM scores impact children's social exclusion in complex moral scenarios, but that this ability does not provide any advantage in a simpler and more straightforward scenario, such as the prototypic moral transgression in the intentional unfair advantage context.

General Summary. The analyses reveal that children's MoToM performance significantly mediated the effects of group membership on children's person judgments for two of three advantage contexts. As predicted, MoToM was a meaningful mediator for both of the morally complex scenarios (unintentional unfair advantage and intentional fair advantage) but did not successfully mediate the relationship in the morally straightforward scenario (intentional unfair advantage). This supports the idea that children's intergroup biases can potentially be mediated by their MoToM competence, but that this mechanism is only useful in scenarios when biases are activated and the intentions of an individual are complex, such as in the unintentional unfair and intentional fair contexts. The intentional unfair advantage was a straightforward moral transgression which most children viewed as wrong. The ability for MoToM to serve as a meaningful buffer against children's ingroup biases in these contexts was made more interesting by the finding that fewer children displayed MoToM competence when asked to evaluate the intentions of an outgroup member (H1).

Discussion

In their daily lives, children's understanding of others' mental states is necessary not only for cognitive development but for social and moral development as well (Carpendale & Lewis, 2004; Lagattuta & Weller, 2014). Children are immersed in situations where an understanding of the mental states of others are central to their interpretations of others' intentions in their peer relationships. The current study was designed to assess children's cognitive ability to recognize the mental states of others in social and moral contexts and to document how children's ability to infer the mental states of others was related to their group biases regarding person judgments (attributions of intentions and social exclusion). This study examined the role of group biases in children's mental state understanding, showing that group biases do impact children's abilities to

infer other's mental states. While previous work has demonstrated that children's morally-relevant Theory of Mind (MoToM) skills serve as a predictor in their moral evaluations (D'Esterre et al., 2019; Fu et al., 2014; Li et al., 2017), the current study extended previous work by showing that MoToM competence was a significant mediator between children's group membership and person judgments. Further, the relationship varied depending on contexts varying in underlying intentions and moral outcomes.

In complex contexts where positive intentions resulted in an advantage for one team over another (unintentional unfair advantage and intentional fair advantage) children's MoToM skills mediated the relationship between their group membership and person judgments. However, when the context was simpler and intentions and outcomes were more closely aligned, in a straightforward moral transgression (intentional unfair advantage), children's MoToM skills no longer served as a mediator. This study demonstrated the complex interaction between children's MoToM competence and group membership when children were making two types of person judgments: attributions of intentions and social exclusion. No study, that we know of, has investigated these questions, and as such this study had several novel findings.

Group Membership Predicts MoToM Competence

The relationship between group membership and children's ability to understand characters' mental states was investigated. It was found that children who were evaluating mental states of an ingroup member were more accurate than children evaluating mental states of an outgroup member. Previous research has shown that a shared group membership impacted children's ToM abilities with real social groups (e.g., race/ethnicity) (Gönültaş et al., 2019; McLoughlin & Over, 2017). This study extended those findings to include novel, minimal groups. Even in a minimal group context, children's group membership significantly impacted

their ability to consider outgroup members' mental states. This finding has significant implications for researchers interested in intergroup cognition and intergroup relations, as it suggests that the perception of another individual as an ingroup member or outgroup member has the potential to meaningfully change whether children will accurately interpret their mental states.

One possible explanation for this finding is a motivational explanation: in a competitive context, children are motivated to take the perspective of ingroup members, and thus perform better on the MoToM questions. In this case, children on the blue team shared a group membership with the advantage creator and were motivated to consider alternative reasons and explanations for why the advantage creator acted the way s/he did. On the other hand, children who were on the red team did not share a group membership with the advantage creator and were not motivated to consider underlying motives or alternative explanations—they only needed to know that a rule was broken by an outgroup member in order to make their evaluations. Thus, children on the two teams had different underlying motivations for considering the mental state of the advantage creator, perhaps explaining the difference in MoToM performance by group. This is just one possible explanation for the different MoToM abilities for children in the ingroup and outgroup conditions, and more research is needed to investigate the underlying mechanisms for these judgments.

One limitation of the study is that the direction of the effect could not be determined: namely if participants in the ingroup condition received a "boost" to their MoToM performance or whether those in the outgroup condition had their MoToM abilities inhibited, or whether both processes were at play. One way in which this relationship could be determined is through the addition of a neutral reference condition which is not assigned to either team and would serve as

a comparison point for the ingroup and outgroup participants. Another way to determine the mechanism would be to ask participants for their reasoning, which would further clarify participants decisions. It is also possible that extreme identification with the ingroup or high levels of negativity towards the outgroup could impact both the "boosting" or inhibiting of MoToM abilities. Future work should consider the underlying mechanism by which group membership impacts children's mental state understanding and how the mechanism plays out (e.g., boosting, inhibiting, or both).

Additionally, future work should consider how various ToM measures may contribute different or additional information to the understanding of intergroup mental state understanding. The present study utilized MoToM over other measures of ToM (i.e. second order ToM, prototypic ToM) because of its ecological validity and applicability in competitive intergroup contexts with social and moral information. However, future research should consider differences in information collected from each type of task and their specific applicability across contexts and study designs.

Attributions of Intentions

The finding that children have better MoToM skills when considering an ingroup member is interesting when it is considered alongside the second main finding of this study: children's MoToM competence mediated the relationship between their ingroup biases and their AoI to the advantage creator. Specifically, children who displayed advanced MoToM competence were less likely to display an ingroup bias when assessing the intentions of others or deciding if a character should be excluded from a team activity. This suggests that children who do not have strong mental state understanding fall back on group membership when making judgments about intentions. In contrast, children who have strong mental state understanding do

not rely as heavily on group membership when making their judgments about intentions. Importantly, researchers investigating the numerous ways in which ingroup biases can impact children's intergroup evaluations and social decisions should consider how mental state understanding may also be affecting the relations of interest.

This study found significant and expected differences in children's AoI based on group membership: children who shared a group membership with an actor evaluated their intentions more positively than children who did not share a group membership. This supports previous work showing that children evaluate ingroup members more positively than outgroup members (McAuliffe & Dunham, 2016; Nesdale et al., 2005). The findings also extend previous work by showing that children's cognitive abilities to infer others mental states can mediate the ingroup bias process. Children who performed better on the MoToM task showed smaller ingroup biases than those who displayed less MoToM competence. While MoToM skills can serve an instrumental role in reducing ingroup bias, they are also impacted by ingroup bias. Future research should further investigate when MoToM skills are both impacted by and can reduce the effects of ingroup biases.

The use of the AoI measure extends the literature and provides further evidence that even young children are able to perform this complex person judgment. Previous work examining attributions of intentions showed that young children are able to distinguish between good and bad intentional and unintentional claims to resources (D'Esterre et al., 2019). This study adds to that literature by examining attributions of intentions across three intergroup contexts that varied in intentional and moral outcomes. Children were able to distinguish between the varying intentional and moral outcomes and attribute intentions differentially depending on the contexts. While previous work documented that children are able to attribute intentions in complex moral

scenarios (D'Esterre et al., 2019), the current work shows that this ability is further impacted by both group processes (ingroup bias) and context. Future work using the AoI measure should be careful to consider how contextual factors, as well as cognitive factors (e.g., ingroup biases, mental state understanding) might impact the findings. This relatively new line of research offers new possibilities for measuring children's person judgments in new and interesting ways.

Additionally, to the best of our knowledge the mental state question utilized for our MoToM measure is the only morally-relevant form of theory of mind that has been utilized in developmental research. However, it is possible that, similar to the standard theory of mind scale (Wellman & Liu, 2004), there may be multiple aspects of MoToM. Future research should investigate other aspects of MoToM and its ability to provide insights into children's intergroup interactions and evaluations.

Future research into the various types of MoToM abilities that children possess, the developmental trajectory of each, and the social and cognitive factors which are related to their development and expression are all fruitful avenues of future research. It is also worth noting that the valence of a character's intentions is only one possible person judgment out of a robust body of literature focused on group biases and person judgments. Further research on these topics may benefit from using a similar framework to investigate the role of group identity and intention understanding on children's judgments of numerous traits (e.g., kindness, trustworthiness, etc.).

Social Exclusion

MoToM competence was not only important for mediating the relationship between group membership and children's AoI, it also served as a mediator between children's group membership and decisions to exclude a peer. One goal of the study was to determine whether MoToM competence functioned as a mediator when the outcome variable was less directly

related to intentions, such as children's decisions to socially exclude a peer. Here, significant and expected differences were found such that children's mental state understanding predicted their social exclusion decisions: children who shared a group membership with the advantage creator were more likely to have higher mental state understanding and less likely to exclude that character than children who did not share a group membership with them. This lends additional validity to the idea that MoToM skills can mediate the role of children's ingroup biases on their person judgments and extends the generalizability of the concept. This finding extends previous studies showing that children's mental state understanding is predictive of their social exclusion decisions (Abrams et al., 2009, 2014) by examining the relationship between MoToM and social exclusion of an ingroup or outgroup team member.

Future work should also consider how these processes play out when children are making *social inclusion* decisions. Previous work has shown that asking children and adolescents to predict inclusion of peers can reveal biases and stereotypes that are not apparent when asking participants to predict exclusion of peers (Hitti & Killen, 2015). While the current study only asked for social exclusion predictions, it is possible the results are even stronger and could reveal more underlying biases if children predicted inclusion decisions. The more subtle nature of asking children to predict social inclusion allows children to answer positively and negatively, while focusing solely on exclusion limits the valence of the question to being solely negative (Møller & Tenenbaum, 2011). Thus, future work may be able to reveal more detailed nuances in the complex relationship between group membership, mental state understanding, and social decisions by including questions about peer inclusion.

Social Context Impacts these Complex Relationships

Interactions between group identity, mental state understanding, and moral cognition are most likely to occur in scenarios where all three elements are present and highly salient. Previous research has indicated that competition increases the importance of group identity and influences children's resource allocation behavior (McGuire et al., 2017). Misunderstandings in competitive contexts are something that is ecologically valid and familiar to children. These competitive intergroup contexts are also the natural intersection of group identity, moral cognition, and mental state understanding. Previous work has shown that children with more advanced mental state understanding were more likely to expect others to challenge group based cognition (Mulvey et al., 2016) and therefore there is reason to believe that mental state understanding could also serve as a buffer against the effects of ingroup favoritism or bias in these high-stakes competitive scenarios. Given that a competitive context seems to heighten concerns for group affiliation and can influence children's fairness decisions this context provides an excellent starting point for investigating the potential mediation of mental state understanding on group identity and person judgments.

Importantly, the relation between MoToM and intergroup biases was not found for each of the contexts that children evaluated. MoToM competence was a mediator of ingroup bias for the unintentional unfair and intentional fair contexts, but not for the intentional unfair context. This is consistent with previous research on children's moral judgments which show that while children may differ in their evaluation of complex situations, straightforward prototypic transgressions are universally seen as negative (D'Esterre et al., 2019; Smetana et al., 2014).

This study sheds light on the cognitive processes where ingroup bias can and cannot impact children's person judgments. Specifically, a competitive context alone is not enough to create reliable ingroup biases. When children witnessed a straightforward moral transgression,

the majority of children evaluated intentions accurately and were more willing to exclude. In this case, fairness overrides children's ingroup biases. The fact that MoToM competence does not serve as a mediator between ingroup bias and person judgments in the straightforward moral transgression, suggests that there may be other contexts where MoToM competence may not help children with their judgments. Future studies should consider when and why children may rely on MoToM competence.

Additionally, it is worth noting that a fourth possible combination of intentions and fairness is possible and was not tested in this study – namely an "Unintentional Fair Advantage" context. This decision not to include the fourth condition was made intentionally as pilot testing revealed that the younger children in our sample struggled to remain engaged when asked to evaluate four conditions. Given that we had stronger theoretical beliefs regarding the other three conditions the Unintentional Fair Advantage was removed from the current study. However, this remains an interesting context for future research and should be considered going forward.

Conclusions

In sum, mental state understanding plays a complicated and critical role in the relationship between children's ingroup biases and their person judgments. Children with MoToM competence are able to accurately attribute intentions of a character and are less likely to exclude this character, regardless of the child's own group affiliation. Yet, this process varies by context such that MoToM was a meaningful mediator for both of the morally complex scenarios (unintentional unfair advantage and intentional fair advantage) but did not mediate the relationship in the morally straightforward scenario (intentional unfair advantage). This is in line with the idea that biases are more prevalent in contexts which are complex (McGlothlin & Killen, 2010) and that children's intergroup biases can potentially be mediated by their MoToM

competence in contexts where those biases are more likely to be revealed. These findings support the propositions put forward by Carpendale and Lewis (2004) that children's developing understanding of the mind occurs within social interactions involving children's experience of the social world and moral knowledge. Mental state understanding is not an individualistic accomplishment but is constructed through social interaction and experience.

In addition, children were better at recognizing and correctly identifying the beliefs of their ingroup member than were children who were asked to attribute intentions and beliefs of an outgroup member. Further supporting Capendale and Lewis' (2004) claim, this suggests that children's mental state understanding is actively and fluidly impacted by context and social processes, such as ingroup bias. This interesting finding leads to many more questions concerning how ingroup biases impact children's, and adults', abilities to accurately take mental perspectives of others.

Developmental scientists moving forward are tasked with considering the complex role of mental state understanding in children's daily person judgments. While it is clear that mental state understanding impacts other relations, such as the relation between ingroup bias and person judgments, it is also the case that mental state understanding is impacted by factors, such as ingroup bias and context. Thus, future work should continue to disentangle the complex relations between mental state understanding, ingroup biases, person judgments, and context.

References

Abrams, D., Rutland, A., & Cameron, L. (2003). The development of subjective group dynamics: Children's judgments of normative and deviant in-group and out-group individuals. *Child Development*, *74*(6), 1840–1856. doi: 10.1046/j.1467-8624.2003.00641.x

Abrams, D., Rutland, A., Palmer, S. B., Pelletier, J., Ferrell, J., & Lee, S. (2014). The role of cognitive abilities in children's inferences about social atypicality and peer exclusion and inclusion in intergroup contexts. *British Journal of Developmental Psychology*, *32*(3), 233–247. doi: 10.1111/bjdp.12034

Abrams, D., Rutland, A., Pelletier, J., & Ferrell, J. M. (2009). Children's group nous: Understanding and applying peer exclusion within and between groups. *Child Development*, *80*(1), 224–243. doi: 10.1111/j.1467-8624.2008.01256.x

Carpendale, J., & Lewis, C. (2004). Constructing an understanding of mind: The development of children's social understanding within social interaction. *Behavioral and Brain Sciences*, *27*(01). doi: 10.1017/S0140525X04000032

D'Esterre, A. P., Rizzo, M. T., & Killen, M. (2019). Unintentional and intentional falsehoods: The role of morally relevant theory of mind. *Journal of Experimental Child Psychology*, *177*, 53–69. doi: 10.1016/j.jecp.2018.07.013

Dunham, Y., Baron, A. S., & Carey, S. (2011). Consequences of "minimal" group affiliations in children. *Child Development*, *82*(3), 793–811. doi: 10.1111/j.1467-8624.2011.01577.x

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. doi: 10.3758/BRM.41.4.1149

Fu, G., Xiao, W. S., Killen, M., & Lee, K. (2014). Moral judgment and its relation to second-order theory of mind. *Developmental Psychology*, *50*(8), 2085–2092. doi: 10.1037/a0037077

Gönültaş, S., Selçuk, B., Slaughter, V., Hunter, J. A., & Ruffman, T. (2019). The capricious nature of Theory of Mind: Does mental state understanding depend on the characteristics of the target? *Child Development*. doi: 10.1111/cdev.13223

Griffiths, J. A., & Nesdale, D. (2006). In-group and out-group attitudes of ethnic majority and minority children. *International Journal of Intercultural Relations*, *30*(6), 735–749. doi: 10.1016/j.ijintrel.2006.05.001

Hitti, A., & Killen, M. (2015). Expectations about ethnic peer group inclusivity: The role of shared interests, group Norms, and stereotypes. *Child Development*, *86*(5), 1522–1537. doi: 10.1111/cdev.12393

Hughes, C., & Devine, R. T. (2015). Individual differences in Theory of Mind from preschool to adolescence: Achievements and directions. *Child Development Perspectives*, *9*(3), 149–153. doi: 10.1111/cdep.12124

Killen, M., & Cooley, S. (2014). Morality, exclusion, and prejudice. In M. Killen & J. G. Smetana (Eds.), *Handbook of Moral Development*, 2nd edition (pp.340-360). NY: Psychology Press. doi: 10.4324/9780203581957.ch16

Killen, M., Mulvey, K. L., Richardson, C., Jampol, N., & Woodward, A. (2011). The accidental transgressor: Morally-relevant theory of mind. *Cognition*, *119*(2), 197–215. doi: 10.1016/j.cognition.2011.01.006

Killen, M., Pisacane, K., Lee-Kim, J., & Ardila-Rey, A. (2001). Fairness or stereotypes? Young children's priorities when evaluating group exclusion and inclusion. *Developmental Psychology*, *37*(5), 587–596. doi: 10.1037/0012-1649.37.5.587

Lagattuta, K. H. (2005). When you shouldn't do what you want to do: Young children's understanding of desires, rules, and emotions. *Child Development*, *76*(3), 713–733. doi: 10.1111/j.1467-8624.2005.00873.x

Lagattuta, K., & Weller, D. (2014). Interrelations between theory of mind and morality: A developmental perspective. In M. Killen & J. Smetana (Eds.), *Handbook of moral development* (pp. 385–408). Psychology Press.

Leslie, A. M., Knobe, J., & Cohen, A. (2006). Acting intentionally and the side-effect effect: Theory of Mind and moral judgment. *Psychological Science*, *17*(5), 421–427. doi: 10.1111/j.1467-9280.2006.01722.x

Li, L., Rizzo, M. T., Burkholder, A. R., & Killen, M. (2017). Theory of mind and resource allocation in the context of hidden inequality. *Cognitive Development*, *43*, 25–36. doi: 10.1016/j.cogdev.2017.02.001

McAuliffe, K., & Dunham, Y. (2016). Group bias in cooperative norm enforcement. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1686), 20150073. doi: 10.1098/rstb.2015.0073

McGlothlin, H., & Killen, M. (2006). Intergroup attitudes of European American children attending ethnically homogeneous schools. *Child Development*, *77*(5), 1375–1386. doi: 10.1111/j.1467-8624.2006.00941.x

McGlothlin, H., & Killen, M. (2010). How social experience is related to children's intergroup attitudes. *European Journal of Social Psychology*, *40*(4), 625–634. doi: 10.1002/ejsp.733

McGuire, L., Manstead, A. S. R., & Rutland, A. (2017). Group norms, intergroup resource allocation, and social reasoning among children and adolescents. *Developmental Psychology*, *53*(12), 2333–2339. doi: 10.1037/dev0000392

McGuire, L., Rutland, A., & Nesdale, D. (2015). Peer group norms and accountability moderate the effect of school norms on children's intergroup attitudes. *Child Development*, *86*(4), 1290–1297. doi: 10.1111/cdev.12388

McLoughlin, N., & Over, H. (2017). Young children are more likely to spontaneously attribute mental states to members of their own group. *Psychological Science*, *28*(10), 1503–1509. doi: 10.1177/0956797617710724

Møller, S. J., & Tenenbaum, H. R. (2011). Danish majority children's reasoning about exclusion based on gender and ethnicity: Children's reasoning about exclusion. *Child Development*, *82*(2), 520–532. doi: 10.1111/j.1467-8624.2010.01568.x

Mulvey, K. L. (2016). Children's reasoning about social exclusion: Balancing many factors. *Child Development Perspectives*, *10*(1), 22–27. doi: 10.1111/cdep.12157

Mulvey, K. L., Buchheister, K., & McGrath, K. (2016). Evaluations of intergroup resource allocations: The role of theory of mind. *Journal of Experimental Child Psychology*, *142*, 203–211. doi: 10.1016/j.jecp.2015.10.002

Mulvey, K. L., Rizzo, M. T., & Killen, M. (2016). Challenging gender stereotypes: Theory of mind and peer group dynamics. *Developmental Science*, *19*(6), 999–1010. doi: 10.1111/desc.12345

Nesdale, D. (2004). Social identity processes and children's ethnic prejudice. In M. Bennett & F. Sani (Eds.), *The development of the social self*. NY: Taylor & Francis. doi: 10.4324/9780203391099

Nesdale, D., Durkin, K., Maass, A., & Griffiths, J. (2005). Threat, group identification, and children's ethnic prejudice. *Social Development*, *14*(2), 189–205. doi: 10.1111/j.1467-9507.2005.00298.x

Nesdale, D., Griffith, J., Durkin, K., & Maass, A. (2005). Empathy, group norms and children's ethnic attitudes. *Journal of Applied Developmental Psychology*, *26*(6), 623–637. doi: 10.1016/j.appdev.2005.08.003

Rizzo, M. T., & Killen, M. (2018). How social status influences our understanding of others' mental states. *Journal of Experimental Child Psychology*, *169*, 30–41. doi: 10.1016/j.jecp.2017.12.008

Rutland, A., & Killen, M. (2015). A developmental science approach to reducing prejudice and social exclusion: Intergroup processes, social-cognitive development, and moral reasoning. *Social Issues and Policy Review*, *9*(1), 121–154. doi: 10.1111/sipr.12012

Smetana, J. G., Jambon, M., & Ball, C. (2014). The social domain approach to children's moral and social judgments. In M. Killen & J. G. Smetana (Eds.), *Handbook of moral development* (2nd ed., pp. 23–45). Psychology Press.

Verkuyten, M. (2007). Social psychology and multiculturalism. *Social and Personality Psychology Compass*, *1*(1), 280–297. doi: 10.1111/j.1751-9004.2007.00011.x

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-Analysis of Theory-of-Mind Development: The truth about false belief. *Child Development*, *72*(3), 655–684. doi: 10.1111/1467-8624.00304

Wellman, H. M., & Liu, D. (2004). Scaling of Theory-of-Mind tasks. *Child Development*, *75*(2), 523–541. doi: 10.1111/j.1467-8624.2004.00691.x