



PAPER

Challenging gender stereotypes: Theory of mind and peer group dynamics

Kelly Lynn Mulvey,¹ Michael T. Rizzo² and Melanie Killen²

1. Department of Educational Studies, University of South Carolina, USA

2. Department of Human Development and Quantitative Methodology, University of Maryland, USA

Abstract

To investigate the social cognitive skills related to challenging gender stereotypes, children (N = 61, 3–6 years) evaluated a peer who challenged gender stereotypic norms held by the peer's group. Participants with false belief theory of mind (FB ToM) competence were more likely than participants who did not have FB ToM to expect a peer to challenge the group's stereotypes and propose that the group engage in a non-stereotypic activity. Further, participants with FB ToM rated challenging the peer group more positively. Participants without FB ToM did not differentiate between their own and the group's evaluation of challenges to the group's stereotypic norms, but those with ToM competence asserted that they would be more supportive of challenging the group norm than would the peer group. Results reveal the importance of social-cognitive competencies for recognizing the legitimacy of challenging stereotypes, and for understanding one's own and other group perspectives.

Research highlights

- False belief theory of mind (FB ToM) is related to expecting that peers will challenge group norms about gender stereotypic activities.
- Children with FB ToM are more positive about others who challenge gender stereotypic norms than are children without FB ToM.
- Children without FB ToM do not differentiate between their own judgments, and their predictions about how the group will respond.

Introduction

Gender stereotypes emerge during the preschool years and are deeply entrenched by adulthood (Eagly & Wood, 2013). Much research has examined the factors that contribute to gender-related expectations; however, less is known about what variables constitute the obstacles

for challenging gender stereotypes. For example, while young children recognize and identify gender stereotypes (Ruble, Martin & Berenbaum, 2006), children also reject social exclusion based on gender stereotypes (Killen, Pisacane, Lee-Kim & Ardila-Rey, 2001). What may be more difficult, though, is supporting a peer who challenges one's own group to consider engaging in cross-gender activities, a form of intragroup deviation. Researchers have called for more work that examines intragroup deviation, which involves rejecting the norms of one's own group (Nesdale, Lawson, Durkin & Duffy, 2010). We propose that engaging in such a challenge may hinge upon theory of mind (ToM) competence, that is, the ability to recognize that others may have different beliefs, desires and intentions from oneself (Wellman & Liu, 2004). This is because mental state knowledge, such as false belief (FB) ToM, provides a social-cognitive basis for challenging group norms about stereotypes, as we elaborate below.

To know that a group may not be favorable to change their norms, even when an individual member wants the change, requires coordinating two perspectives about the

Address for correspondence: Kelly Lynn Mulvey, 129 Wardlaw, Department of Educational Studies, University of South Carolina, Columbia, SC 29208, USA; e-mail: mulveykl@mailbox.sc.edu

same set of beliefs. In the case of intragroup deviation regarding gender stereotypic activities, for example, one has to know that the group has a belief that their group members like to play with dolls, and, at the same time, that one member of the group does not like to play with dolls (the deviant). Thus, when children consider challenging group norms or evaluate others' decisions to challenge such norms (e.g., by suggesting that the group act differently), they have to weigh two perspectives: the group's perspective and the individual's viewpoint (Abrams & Rutland, 2008; Mulvey, Hitti, Rutland, Abrams & Killen, 2014; Nesdale & Lawson, 2011). The ability to anticipate others' responses to a deviant member could lead to a greater willingness to challenge social-group norms to affect a change in others' attitudes. The ability to predict responses, however, may also hinder children's resistance, as they may expect that their group will not support such a challenge, leading to compliance. Importantly, in the current study we will measure both participants' predictions about their peers' likelihood of resistance as well as their own evaluations of how likely they would be to resist. We included both measures as children, especially those with more sophisticated social-cognitive skills, may recognize that others may not support a group's norm, even if the child individually does support such a norm or vice versa.

Research has not fully examined children's willingness to challenge peer group norms, which reflect larger societal patterns, customs, and even stereotypes. Group norms exert a powerful influence on the behavior and judgments of children (Abrams & Rutland, 2008). By 6 to 8 years of age, children are attuned to group dynamics and value someone who is not a group member but holds the group's norms. While some research has examined the relation between social-cognitive skills, such as classification skill, and gender stereotyping (Bigler, 1995), less is known about the social-cognitive basis for challenging stereotypic group norms. No research, to our knowledge, has directly examined challenges to peer group norms as related to their mental state understanding, referred to as ToM competence.

Children's mental state knowledge develops rapidly during the preschool period. While research indicates that some forms of intentionality understanding may be present as early as infancy (see Carlson, Koenig & Harms, 2013, for a review), children exhibit increasing skill with explicit ToM tasks between the ages of 3 and 6 years (Wellman & Liu, 2004). We propose that false belief theory of mind (FB ToM) may be especially relevant to situations involving challenging peer norms. FB ToM, which typically develops between the ages of 3 and 5 years, involves recognizing that an actor may hold

a 'false belief' or inaccurate understanding of a situation (Wellman & Liu, 2004). Research on children's mental state understanding reveals that children with FB ToM make more sophisticated moral judgments than those without FB ToM (Killen, Mulvey, Richardson, Jampol & Woodward, 2011; Smetana, Jambon, Conry-Murray & Sturge-Apple, 2012). Research has also shown that children with strong FB ToM skills are more persuasive (Slaughter, Peterson & Moore, 2013) and have stronger social skills than those without FB ToM (Watson, Nixon, Wilson & Capage, 1999). Thus, FB ToM competence plays an important role in children's ability to understand and engage in multiple dimensions of peer interactions.

We expect that FB ToM competence will be related to children's likelihood of challenging stereotypic group norms. This is because FB ToM enables children to recognize that others can have different mental states (beliefs, desires, emotions, etc.) from them. Children with FB ToM, then, have the necessary cognitive competencies to recognize that when a group norm is predicated upon a stereotype, not all members of the group may share the same beliefs: some members may recognize that relying upon stereotypes could lead to differential or unfair treatment of others. Thus, children with greater FB ToM competence, who can distinguish their own perspective from a group perspective, should be more willing to challenge stereotype-based group norms, and support those who do so as well. We measured both individual resistance to group norms as well as expectations about peer resistance to group norms. Including both measures provides greater insight into group dynamics, allowing us to examine separately expectations about and support for peer deviants, following in the tradition of subjective group dynamics (Abrams & Rutland, 2008), as well as individual expectations.

Current study

The current study examined whether children with and without FB ToM would support challenges to gender stereotypic norms regarding their choice of activities and their expectations of peer group responses to such challenges. The study draws upon social domain theory (Smetana, Jambon & Ball, 2014; Turiel, 1983) in examining children's social and moral reasoning, and developmental intergroup theory for challenging group norms (Killen & Rutland, 2011). The social domain model identifies three domains of social reasoning, the moral domain (justice and welfare), the societal domain (customs, traditions and group norms), and the psychological domain (personal choice, psychological knowledge and autonomy). Recent extensions to social domain

theory have proposed that children struggle with social decisions when group norms conflict with individual norms or values (Mulvey, Hitti & Killen, 2013). In this study, all three domains play important roles: children may consider the moral domain (for instance, recognizing that stereotypes can lead to biased, prejudicial behavior that harms others), the societal domain (for instance, perceiving the pressure to show loyalty to the group and maintain the group norms) and the psychological domain (for instance, applying psychological knowledge and recognizing that individual beliefs may be different from group norms).

The current study focused on gender stereotypic norms, as gender stereotypes are widely held among young children (Liben & Bigler, 2002; Ruble *et al.*, 2006) and gender is an intergroup category that emerges quite early (Martin & Ruble, 2004). Children use information about gender stereotypes when making predictions about toy choices their peers might make (Martin, Eisenbud & Rose, 1995). Further, some research already supports the idea that children with ToM may be more open-minded about gender norms than those without ToM; children with FB ToM are less likely to assume gender discrimination in ambiguous scenarios (Brown, Bigler & Chu, 2010). Thus, focusing on how ToM abilities influence children's understanding of group dynamics surrounding gender stereotypes is an ideal candidate for research.

Further, gender stereotypes may lead to false predictions about another's interests, desires or behaviors. This is because while gender stereotypes can often accurately reflect others' desires (for instance it is true that many boys enjoy or even prefer playing with trucks), there are also many instances when such stereotypes are inaccurate and do not reflect others' interests or desires. It is not the case that all girls prefer playing with dolls or all boys prefer trucks. Thus, challenging gender stereotypes may require the ability to recognize variation in perspectives and to be able to perceive that not all group members or peers will support a group's gender stereotypic norms. Thus, it may be a false prediction to decide that any given girl wants to play with dolls just because she is a girl (when this is not necessarily the case given that some girls do not like playing with dolls); the same is true for boys playing with trucks.

The study included participants between 3 and 6 years of age, as during this period children develop experience in groups and acquire more sophisticated FB ToM skills (Wellman & Liu, 2004). Participants' awareness of gender stereotypes and their FB ToM were assessed. We measured children's expectations for whether group members would challenge their group and their perceptions of the group response. Participants evaluated two

types of stories. In one, participants heard about a non-conforming group member who wanted the group to play with a gender non-stereotypic toy (as determined by prior research on gender stereotypic knowledge; see Liben & Bigler, 2002). In this story, the non-conforming group member and all other members had access to the same information about the toy choices available; thus this was called the *Belief: One Perspective* condition. In the *Belief: Two Perspectives* condition, the group held the false belief that their preferred toy (gender stereotypic) was available, while the non-conforming member knew that the toy was broken. Both types of stories were included as prior research has shown that children with FB ToM responded differently from children without FB ToM to morally relevant scenarios which did and did not involve explicit false beliefs (Killen *et al.*, 2011). One aim was to test whether this same pattern is present in contexts that involve social norms, such as those which govern activity choice. Given prior research which indicated that FB ToM was related to judgments in contexts which did and did not explicitly involve false beliefs (Killen *et al.*, 2011), we expected that participants with FB ToM would show greater sophistication than participants without FB ToM in both the *Belief: One Perspective* and in the *Belief: Two Perspectives* conditions.

We expected that (1) participants with FB ToM would be more likely to support challenging the group's gender stereotypic norm, and would be more likely to expect others to do so, in response to stories which did and did not involve explicit false beliefs. Children with greater FB ToM competence are more socially engaged, and may be more likely than children without FB ToM competence to express their opinion and negotiate a resolution if they disagree (Rochat & Ferreira, 2008). Further, children with FB ToM may be more open-minded about gender roles (Brown *et al.*, 2010) and may be more flexible in their thinking about gender stereotypes (Martin & Ruble, 2004; Trautner, Ruble, Cyphers, Kirsten, Behrendt *et al.*, 2005).

Based on past findings (Mulvey *et al.*, 2014), it was also expected that (2) participants without FB ToM would be unable to distinguish their own perspective from the group's perspective, but that participants with FB ToM would understand that the group would evaluate a non-conforming act differently than they would. Finally, it was expected that (3) children's social reasoning would differ depending on whether they supported a non-conforming group member or not. For instance, we expected that participants who were less favorable towards the non-conforming member would reference gender stereotypes more often than children who were more favorable.

Given the lack of previous findings regarding FB ToM and social skills based on gender (Killen *et al.*, 2011), it was an open question whether participant gender would be related to children's judgments. On the one hand, participant gender could play a role in that children would be less willing to support challenging their own gender group than the opposite gender's group, even when participants had FB ToM, due to ingroup preferences. On the other hand, social cognitive skills or gender associations with activities could override gender ingroup bias in the context of challenging group norms. For these reasons, participant gender differences were tested despite the absence of specific gender-related hypotheses. Further, prior research has found that acting in non-stereotypic ways is often seen as more acceptable for girls than for boys (Horn, 2007, 2008; Smetana, 1986; Zucker, Wilson-Smith, Kurita & Stern, 1995). The current study, however, involves a less overt form of non-stereotypic behavior: activity choice. Therefore, it was an open question whether evaluations would vary based on gender of the target (girls' group, boys' group) or whether participant gender would interact with gender of the target. In sum, we tested for differences based both on participant gender and the gender of the target in the story; however, we did not have specific hypotheses as prior research could reasonably indicate that we may or may not find differences based on participant gender or the gender of the target group in the stories.

Materials and methods

Participants

Participants included children 3 to 6 years of age ($N = 61$; $M_{age} = 57.7$ months), approximately evenly divided by gender (45% female), from a preschool in the mid-Atlantic region of the United States, and with middle income backgrounds and approximately 30% ethnic minority participants (10% Latino, 15% Asian, 5% African-American) with 70% ethnic majority participants. Informed parental consent was obtained for all participants.

Design and procedures

Participants first completed warm-up tasks to familiarize them with the Likert-type scales that were to be used as has been employed in previous studies (see Cooley & Killen, 2015; Killen *et al.*, 2011). Participants completed unrelated sample items ('How much do you like pizza?')

indicating their choices on 6-point Likert-type scale. They were familiarized with two scales; the experimenter provided examples of how to use the middle of the scale (e.g., 'I only like pizza a little because sometimes it burns my mouth so I would point to 4'). One included faces which ranged from a frowning face to a smiling face and was used to measure how okay or not okay participants judged items to be. One included the words 'Yes' and 'No' in increasingly larger font size and was used to measure how likely or not likely participants judged items to be. Participants completed the *Challenging the Peer Group Task*, which included four stories. Two stories (about a boys' group and a girls' group) were *Belief: One Perspective* stories, where a single member of the group wanted to challenge the group norm by playing with a non-stereotypic toy. An example of a *Belief: One Perspective* story for a male participant was the following:

Your group, the boys' group likes to play with the race cars. During playtime, your group, the boys' group says 'Let's play with the race cars, those are for boys.' Today, Frank, in your group, the boys' group, wants to be different from his group. He doesn't want to play with the race cars. He wants to say 'People think tea sets are only for girls, let's play with the tea set.'

The other two were *Belief: Two Perspectives* stories, where a single member of the group wants to challenge the group to play with a non-stereotypic toy. In the *Belief: Two Perspectives* story, the non-conforming member, but not the rest of the group, knows that the gender stereotypic toy is broken, thus the group holds one perspective on the status of the toys, while the target holds a different perspective. An example of the *Belief: Two Perspectives* story for a male participant was the following:

Your group, the boys' group also likes to play with the fire house. Dave, in your group, the boys' group, gets to school before the rest of the boys group. He sees the toy box open and notices that the fire house is broken. Before the other children arrive, the teacher closes up the toy box. Later, during playtime, your group, the boys' group says 'Let's play with the fire house.' ... Today, Dave, in the boys group, wants to be different from his group. He doesn't want to play with the fire house. He knows it is broken. He wants to say 'People think doll houses are only for girls, let's play with the doll house.'

The groups were described as either 'your group' or 'their group' depending on the participant's gender and all participants evaluated stories for both boys' and girls' groups. Participants also completed the *Gender*

Association task, which measured participants' attitudes towards typical gender associations, and the *False Belief ToM* task, which measured participants' FB ToM ability. The tasks were administered in the school by a trained researcher. The assessment took approximately 25–35 minutes to complete.

Measures

Challenging Gender Stereotypes task

The dependent measures for the *Challenging Gender Stereotypes task* are displayed in Table 1.

Coding categories for justifications

A coding system was established drawing on prior research from social domain theory (Killen *et al.*, 2001; Killen, Rutland, Abrams, Mulvey & Hitti, 2013). The coding categories and example codes are displayed in Table 2. The subcategories for Social-Conventional were: Group Functioning, and Gender Stereotypes. The subcategories for Psychological were: Autonomy, and Personal Identification. Moral reasoning was also coded, but was used less than 10% of the time and, thus, was not analyzed. Reasoning was identified as Uncode-

Table 1 Measures for the Challenging Gender Stereotypes task

Measure	Question	Response
Expectations regarding peer likelihood of resistance	Do you think she will tell the girls' group that she wants to X?	Likert-type: 1 = Really Won't Tell to 6 = Really Will Tell
Individual likelihood of resistance	What if it was you? Would you tell them?	Likert-type: 1 = Really Won't Tell to 6 = Really Will Tell
Group act evaluation	Let's say Sarah tells the girls' group X. How okay or not okay will they think that is?	Likert-type: 1 = Really Not Okay to 6 = Really Okay
Individual act evaluation Individual act evaluation: reasoning	Now, what about you? When you hear her, how okay or not okay do you think that is? Why?	Likert-type: 1 = Really Not Okay to 6 = Really Okay Open-ended: Why?
Individual act evaluation: reasoning	Why?	Open-ended: Why?

able if it was undifferentiated or incomplete. Reasoning was coded as 1 = full use of the category; .5 = partial use; 0 = no use, and analyses were conducted on proportional usage. Less than 5% of participants used more than one code. Coding was conducted by two unique coders on 25% of interviews and inter-rater reliability was Cohen's $\kappa = .85$.

Gender Association task

The *Gender Association* task measured participants' attitudes towards the gender associations for each of the toy items used in the stories (stereotypically male toys: trucks, tools; stereotypically female toys: tea set, doll house), assessing (1) Gender Association Attitudes: Who should play with X? (boy, girl, or both). Responses were scored as a 0 if they chose the non-stereotypic gender or both, and 1 if they chose the stereotypic gender. Scores were summed across all four questions and responses were split into a low (0–2) and high stereotypic attitudes (3–4) dichotomous variable.

False Belief Theory of Mind task

The wording of the *False Contents ToM* task was:

See this box (pointing to a crayon box). This is a crayon box. Now here is Marta [or Mark], she [or he] is cleaning up the classroom and puts some crackers in the empty crayon box.

The target question was: (1) *Contents False Belief* ('When the other children come back in from playing outside, what will they think is in the crayon box?').

Table 2 Coding categories

Code	Description	Example item
Group functioning	References to ensuring that the group works smoothly, and that group norms are upheld.	'That's what the group likes to play.'
Gender stereotypes	References to explicit gender stereotypes.	'Boys can't play with tea sets.'
Autonomy	References to personal choice.	'It is up to her what she wants to play.'
Personal identification	References to the participants' individual preferences or affiliations.	'I like the doll house, too.'

Participants who responded 'crayons' were scored as having false contents ToM and those who responded 'crackers' were scored as not having ToM. Participants were also given two memory check questions. Participants who failed this memory check were excluded from analyses that included ToM as a variable ($n = 6$ children).

Results

Analyses were conducted and hypotheses were tested using Univariate ANOVAs in order to assess differences based on FB ToM (pass, fail). Analyses included age of participant as a covariate in order to document differences above and beyond age. Reasoning analyses and comparisons across question were conducted as repeated measures ANOVAs. Repeated measures designs are effectively analyzed using ANOVAs because ANOVAs are robust to the problem of empty cells, whereas other data analytic procedures (e.g. log-linear models) require cumbersome data manipulation to adjust for empty cells (see Posada & Wainryb, 2008, for a fuller explanation and justification of this data analytic approach). No differences were found for participant gender, gender association attitudes or between responses to stories about girls' groups and boys' groups. Thus, participant responses were averaged across girls' and boys' groups for both the *Belief: One Perspective* and *Belief: Two Perspectives* stories separately, and participant gender and gender association attitudes were dropped from analyses.

Descriptive statistics

Descriptive statistics were computed to identify how many participants showed FB ToM competence. Results indicated that fewer 3- to 4-year-olds (37%) than 5- to 6-

year-olds (63%) showed FB ToM competence, $\chi^2(1, n = 55) = 6.760, p = .009$.

In terms of Gender Association Attitudes, 52% of 3- to 4-year-olds and 48% of 5- to 6-year-olds showed high stereotypic attitudes, indicating no differences for gender associations by age of the participant. Further, 50% of participants without FB ToM competence and 55% of participants with FB ToM showed high stereotypic attitudes.

Expectations regarding peer likelihood of resistance

In order to test whether (1) participants with FB ToM would be more likely to expect their peers to challenge the group than would participants without FB ToM, two separate 2 (ToM: Pass, Fail) univariate ANCOVAs with age as a covariate were conducted, one for the *Belief: Two Perspectives* stories and one for the *Belief: One Perspective* stories. As expected, in both the *Belief: One Perspective* ($F(1, 51) = 11.27, p = .001, \eta_p^2 = .18$) and the *Belief: Two Perspectives* stories ($F(1, 49) = 19.57, p < .001, \eta_p^2 = .29$), participants with FB ToM were more likely to expect that the non-conforming member would challenge the peer group than were participants without FB ToM (see Table 3).

Individual likelihood of resistance

In order to test if (1) participants with FB ToM would be more likely to support challenging the group than would participants without FB ToM, two separate 2 (ToM: Pass, Fail) univariate ANCOVAs with age as a covariate were conducted, one for the *Belief: Two Perspectives* stories and one for the *Belief: One Perspective* stories. In the *Belief: One Perspective* stories, there was no significant difference, $p = .07$: participants with FB ToM did not differ in their likelihood rating from those without FB ToM. However, in the *Belief: Two Perspectives*

Table 3 Group and individual assessments of *Belief: One Perspective* and *Beliefs: Two Perspectives* conditions for participants with and without FB ToM

	Belief: One Perspective		Belief: Two Perspectives	
	No FB ToM	FB ToM	No FB ToM	FB ToM
Expectations regarding peer Likelihood of resistance	2.90 (1.65) ^a	4.53 (1.30) ^a	2.92 (1.46) ^b	4.87 (1.24) ^b
Individual likelihood of resistance	3.18 (1.40)	4.18 (1.70)	3.29 (1.39) ^c	4.32 (1.61) ^c
Group act evaluation	2.70 (1.62)	2.66 (1.53) ^d	3.27 (1.82) ^h	3.24 (1.70) ^e
Individual act evaluation	2.84 (1.66) ^g	4.14 (1.45) ^{d,g}	3.11 (1.66) ^h	4.79 (1.36) ^{e,h}

Note: ^{a,b,e} $p < .001$; ^{c,g,h} $p < .01$; ^{c,d} $p < .05$. No FB ToM = participants without false belief theory of mind competence. FB ToM = participants with false belief theory of mind competence. 1 = 'Really won't tell' to 6 = 'Really will tell' for peer and individual likelihood of resistance. 1 = 'Really not okay' to 6 = 'Really okay' for group and individual act evaluation.

stories, participants with FB ToM were more likely to support challenging the peer group than participants without FB ToM ($F(1, 50) = 6.14, p = .02, \eta_p^2 = .10$) (see Table 3).

Individual versus group act evaluation

In order to test the hypothesis that (2) participants with FB ToM would understand that the group would not like the non-conforming member's act even though they, individually, would, but that participants without FB ToM would not differentiate between individual and group act evaluation, two separate 2 (ToM: Pass, Fail) \times 2 (Act Evaluation: Individual, Group) ANCOVAs with age as a covariate and repeated measures on the last factor were conducted, one for the *Belief: Two Perspectives* stories and one for the *Belief: One Perspective* stories. As expected, in both the *Belief: One Perspective* stories ($F(1, 50) = 4.27, p = .04, \eta_p^2 = .07$) and the *Belief: Two Perspectives* stories ($F(1, 48) = 9.65, p = .003, \eta_p^2 = .16$) participants with FB ToM asserted that they would condone the non-conforming act more than the group would, while participants without FB ToM did not distinguish between their own and the group's perspective (see Table 3). Supporting this finding, participants with FB ToM were significantly more positive in their act evaluations than were participants without FB ToM, in both the *Belief: One Perspective* stories ($F(1, 45) = 5.18, p = .028, \eta_p^2 = .10$) and the *Belief: Two Perspectives* Stories ($F(1, 43) = 10.89, p = .002, \eta_p^2 = .20$).

Individual act evaluation reasoning

In order to test the hypothesis that (3) there would be differences in children's reasoning about their individual act evaluation of the non-conforming act, analyses were conducted using a dichotomous variable for participants who judged the act of challenging the peer group as okay and those who judged the act as not okay. This variable was computed using a mid-point split of 3.5. The top three forms of reasoning used in the *Belief: One*

Perspective stories were personal identification, autonomy, and gender stereotypes. The top three forms of reasoning used in the *Belief: Two Perspectives* stories were personal identification, group functioning, and gender stereotypes.

Belief: One Perspective stories

In order to assess differences in reasoning in the *Belief: One Perspective* stories, a 2 (Individual Act Evaluation: okay, not okay) \times 3 (Justification: personal identification, autonomy, gender stereotypes) ANOVA with repeated measures on the last factor was conducted. This revealed that participants differed in their use of reasoning ($F(2, 118) = 8.56, p < .001, \eta_p^2 = 0.12$). Follow-up analyses revealed that participants used primarily gender stereotypes, with a smaller proportion referencing personal identification, and autonomy as shown in Table 4. Participants referenced gender stereotypes significantly more than autonomy. There was also an interaction between use of reasoning and act evaluation ($F(2, 118) = 5.52, p = .005, \eta_p^2 = 0.08$). Participants referenced gender stereotypes significantly more when they did not condone the act than when they judged the act as okay. In addition, participants made no use of autonomy reasoning when they thought that challenging the group was not okay, but cited autonomy when they supported challenging the group. Participants did not differ in their use of personal identification reasoning if they were or were not supportive of challenging the group.

Belief: Two Perspectives stories

In order to assess differences in reasoning about the individual act evaluation in the *Belief: Two Perspectives* stories, a 2 (Individual Act Evaluation: okay, not okay) \times 3 (Justification: personal identification, group functioning, gender stereotypes) ANOVA with repeated measures on the last factor was conducted. This revealed that participants differed in their proportional use of

Table 4 Proportional use of reasoning about individual act evaluation in the *Belief: One Perspective* and *Belief: Two Perspectives* stories

	Belief: One Perspective			Belief: Two Perspectives		
	Okay	Not okay	Total	Okay	Not okay	Total
Gender Stereotypes	.25 (.36) ^b	.53 (.44) ^b	.34 (.41) ^a	.21 (.33) ^f	.60 (.46) ^f	.28 (.37) ^{d,e}
Personal Identification	.23 (.33)	.18 (.29)	.21 (.34)	.20 (.37)	.10 (.21)	.18 (.35) ^d
Autonomy	.00 (.00) ^c	.19 (.36) ^c	.13 (.28) ^a	—	—	—
Group Functioning	—	—	—	.12 (.21)	.10 (.21)	.11 (.21) ^e

Note: ^{a,e} $p < .001$; ^{b,c,d} $p < .05$; ^f $p < .01$.

reasoning ($F(2, 110) = 7.39, p = .001, \eta_p^2 = 0.11$). Participants referenced gender stereotypes more than personal identification, and group functioning (see Table 4). An interaction between use of reasoning and individual act evaluation ($F(2, 110) = 4.87, p = .012, \eta_p^2 = 0.08$) revealed that participants referenced gender stereotypes significantly more when they did not condone challenging the group than when they supported challenging the group. In addition, participants did not differ in their use of group functioning or personal identification reasoning when they did or did not like the act of challenging the group.

Discussion

Our novel findings were that children with mental state knowledge supported challenging gender stereotypic group norms. Children with FB ToM were more likely to support such challenges and to understand the group's reactions to peers who reject group norms than children without FB ToM. Furthermore, children with FB ToM were more likely to expect that their peers would resist gender stereotypic group norms and were also more likely to resist such norms themselves, in some contexts. This research demonstrates the importance of examining intragroup dynamics (deviation within the group), as theorized by Nesdale *et al.* (2010). Children supported challenging the group (intragroup contexts) to resist gender stereotypes, indicating that social-cognitive competence is related to supporting decisions to challenge the norms of the group, and particularly when these norms are viewed as unfair.

FB ToM competence was related to one's evaluations of group dynamics involving gender stereotypes. Participants without FB ToM did not think the group would think differently about the non-conforming act than they did, whereas participants with FB ToM understood that the group's perspective about the non-conforming act might differ from their own. Participants with FB ToM expected that groups would evaluate the challenge negatively, even though they personally disagreed. This is a new direction for research on ToM and mental state knowledge. Previous research has focused on when the absence of ToM competence has been relevant for moral judgments (Killen *et al.*, 2011; Smetana *et al.*, 2012) and social decisions, such as resource allocations (Takagishi, Kameshima, Schug, Koizumi & Yamagishi, 2010), but less is known about how ToM competence influences children's willingness to challenge group norms, particularly regarding gender stereotypic expectations. The present results have important implications as they suggest that ToM competence may be related to a

broader range of evaluations and expectations about social situations than was previously thought. Further, this research suggests that future work should examine the influence of ToM in a wider range of social encounters and dilemmas.

Specifically, participants with FB ToM demonstrated support for peers who challenged gender stereotypes, as reflected in higher individual act evaluation ratings. This was true regardless of whether the group held a false belief about the activities, suggesting that children were not simply demonstrating an awareness of the group's mental state knowledge. Rather, in situations involving multiple perspectives, as well as those involving only one belief-state, participants with FB ToM were more supportive of peers who challenge gender stereotypes than were participants without FB ToM. This reveals that FB ToM competence may enable children to critically evaluate group norms, and indicates that FB ToM matters in situations involving different access to information.

In addition, children's likelihood of resisting gender stereotypic group norms and their expectations about peer likelihood were related to FB ToM. Participants with FB ToM were more likely to expect their peers to challenge the group to play with non-stereotypic toys than were those without FB ToM. This was true in both the *Belief: Two Perspectives* and the *Belief: One Perspective* conditions. Thus, participants with FB ToM were more likely to predict that their peers would challenge group norms if they think differently from the group, which may demonstrate a sophisticated understanding of social dynamics. Further, in the *Belief: Two Perspectives* condition, children with FB ToM were more likely to support challenging the gender stereotypic group norm than were children without FB ToM. In the *Belief: One Perspective* condition, however, children with and without FB ToM did not differ in their likelihood of supporting a challenge to the norm. While in the *Belief: One Perspective* condition, children with and without FB ToM had similar access to the mental state awareness of the characters, in the *Belief: Two Perspectives* condition, children with FB ToM had an additional reason (i.e., knowing that the group did not know that the group-norm-toy was broken) to challenge the stereotypic group norm. This knowledge made it more likely for them to challenge the group than children without FB ToM. Having two reasons for which the target might deviate from the stereotypic group norm may make such a challenge easier; thus children with FB ToM may be more likely to think that they, too, would protest the group norm.

These findings suggest that children with advanced social-cognitive skills recognize that gender stereotypes

may lead to false predictions about others' desires and interests and that those with more advanced social-cognitive skills may be better able to recognize that not all individuals share the same perspective. Thus, children with FB ToM may be better able to perceive variability in adherence to gender stereotypes than would children without FB ToM. Prior research on gender stereotypes has indicated that this type of stereotype knowledge peaks around age 5–6 years (Bem, 1981; Ruble *et al.*, 2006), which is at the end of the developmental period when FB ToM competence emerges. While children may develop the cognitive skills necessary to recognize that not all individuals will hold perspectives which reflect stereotypes, stereotypes are perpetuated and condoned across development, even into adulthood (Eagly & Wood, 2013).

These findings demonstrated that as children's social-cognitive skills mature they are able to recognize the importance of challenging unfair, prejudiced group norms, even though they also understand that groups may not like these challenges to their norms. While the current findings cannot conclusively demonstrate why FB ToM matters, we propose that mental state understanding enables children to understand different perspectives. Knowing and reflecting upon these different perspectives may then motivate a sense of fairness, which provides a basis to counter stereotypes, even with opposition from their peer group. In addition, these findings relate to predictions from developmental subjective group dynamics, which proposes that children do not like peers who challenge the group's norms and that understanding group dynamics is linked to social perspective-taking skills (Abrams, Rutland, Pelletier & Ferrell, 2009).

Further, children with FB ToM may be sensitive to the specific types of norms that groups have, and will support individuals who dissent from unfair, prejudiced norms, such as those based on gender stereotypes. This supports recent work which indicates that children like group members who deviate to advocate for an equal allocation of resources (Cooley & Killen, 2015; Killen *et al.*, 2013). Further, no differences by gender or gender association awareness were documented. This indicates that participants may not have been solely relying upon their awareness of societal-expectations or status differentials regarding gender (Leman, Ahmed & Ozarow, 2005), but rather may also have been considering the value of challenging unacceptable peer norms. This reinforces the idea that participants with FB ToM recognize that multiple perspectives abound and that prioritizing gender stereotypic group norms may lead to prejudicial treatment of others. Future research could test other types of group norms, such as norms about

opportunities and access to resources rather than play activities, to further examine the role of participant gender on how social-cognitive competencies are related to challenging gender group norms.

Children's reasoning provided insight into why children supported peers who challenged the group. Children who thought challenging stereotypes was okay reasoned about this choice by considering issues such as autonomy, group functioning, and their own personal identification with the target; they considered many aspects of the benefits of acting against the group. Children who asserted that they would not like it if someone challenged the group, however, focused narrowly on gender stereotypes. Thus, FB ToM may provide children with a better understanding of others' desires and personal choices, which at times conflict with societal expectations about gender conformity. Research shows that children evaluate gender norms as a matter of personal choice (Conry-Murray, 2013; Conry-Murray & Turiel, 2012); the current study extends these findings by demonstrating that children without FB ToM may think more rigidly about gender stereotypes, viewing them as conventional norms.

The current study has a few limitations that future research should address. First, participants made judgments predicting how they and their peers would react in hypothetical scenarios. While past research has successfully used hypothetical vignettes to assess children's moral and social-cognitive abilities (for a review, see Killen & Smetana, 2015), it is also important for future research to utilize alternative methodologies to extend these findings. In addition, we did not measure children's vocabulary or executive functioning. Given that research suggests that ToM performance may be influenced by language skills and verbal ability (Astington & Baird, 2005) as well as executive functioning (Moses, Carlson & Sabbagh, 2005), future research should aim to test relations using non-verbal tasks or include measures of verbal ability and executive functioning to better understand the complex interplay of multiple cognitive skills in children's decision-making and evaluations.

Further, the current study measures children's resistance to stereotypic norms regarding play preferences, yet children hold gender stereotypes about a wide range of traits, behaviors and occupations (Ruble *et al.*, 2006). Future research should test whether children support resistance to norms regarding different types of stereotypes as well as in contexts beyond gender. Future research should continue to examine the relation between FB ToM and challenging peer groups. While the results indicate that children with FB ToM are capable of recognizing the value of challenging the group, do children with FB ToM initiate such challenges

in different ways from children without FB ToM? Children with FB ToM have been found to make more persuasive arguments (Slaughter *et al.*, 2013), and thus may be more skilled in accomplishing change when challenging their peers.

This research demonstrates that children's social-cognitive abilities, specifically FB ToM, are related to children's likelihood of challenging peer group norms. Identifying the variables which are related to challenging gender stereotypes, such as FB ToM, is important as gender stereotypes can lead to harmful and discriminatory behavior (Horn & Sinno, 2014; Zemore, Fiske & Kim, 2000). Even young children, however, assert that they will challenge stereotypes, and moreover, children with FB ToM support challenging gender stereotypic group norms, even though they understand that groups may not support such challenges. These findings have implications in terms of helping children feel comfortable challenging gender stereotypes and increasing acceptance of non-stereotypic behaviors. Specifically, what we demonstrated is that children, especially those with FB ToM, support challenges to gender stereotypic behavior. This suggests that when a child resists stereotypes or acts in a non-stereotypic manner, that many of their peers will stand behind these decisions. It may be that fear or concern over how others will react to non-stereotypic choices is unfounded in many contexts. Prior research has shown promising results when children are explicitly taught to intervene when they hear sexist or gender stereotypic language (Lamb, Bigler, Liben & Green, 2009). The current findings suggest that teachers and parents should continue to encourage children to discuss the ways in which gender stereotypes are problematic, which will provide a comfortable setting for children to test out such challenges. Finally, as indicated by a growing body of research, FB ToM competence is directly connected to the emergence of social-cognitive judgments pertaining to children's peer interactions. These findings provide additional support for children's sophisticated understanding of peer group dynamics and their social world.

Acknowledgements

The first author was supported by the Elizabeth M. Koppitz Fellowship from the American Psychological Foundation and the Ann G. Wylie Dissertation Fellowship from the University of Maryland. The second author is supported by a National Institute of Child Health and Human Development Pre-doctoral Traineeship (#T32HD007542). We thank Shelby Cooley, Laura Elenbaas, Aline Hitti, and Jee Young Noh for their

feedback on this project. We thank the undergraduate research assistants Chelsea Burrell, Emily Catanuso, Nicole Lang, Chandni Patel, Dmitriy Stasishinn and Naina Vohra for their assistance with this project. We are grateful to the students, parents, and teachers who participated in this study.

References

Abrams, D., & Rutland, A. (2008). The development of subjective group dynamics. In S.R. Levy & M. Killen (Eds.), *Intergroup relations and attitudes in childhood through adulthood* (pp. 47–65). Oxford: Oxford University Press.

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

Astington, J.W., & Baird, J.A. (Eds.) (2005). *Why language matters for theory of mind*. New York: Oxford University Press.

Bem, S.L. (1981). Gender schema theory: A cognitive account of sex typing. *Psychological Review*, **88**, 354–364.

Bigler, R.S. (1995). The role of classification skill in moderating environmental influences on children's gender stereotyping: A study of the functional use of gender in the classroom. *Child Development*, **66**, 1072–1087. doi:10.2307/1131799

Brown, C.S., Bigler, R.S., & Chu, H. (2010). An experimental study of the correlates and consequences of perceiving oneself to be the target of gender discrimination. *Journal of Experimental Child Psychology*, **107**, 100–117. doi:10.1016/j.jecp.2010.04.010

Carlson, S.M., Koenig, M.A., & Harms, M.B. (2013). Theory of mind. *WIREs Cognitive Science*, **4**, 391–402. doi:10.1002/wcs.1232

Conry-Murray, C. (2013). Children's reasoning about gender-atypical preferences in different settings. *Journal of Experimental Child Psychology*, **115**, 210–217. doi:10.1016/j.jecp.2012.09.007

Conry-Murray, C., & Turiel, E. (2012). Jimmy's baby doll and Jenny's truck: Young children's reasoning about gender norms. *Child Development*, **83**, 146–158. doi:10.1111/j.1467-8624.2011.01696.x

Cooley, S., & Killen, M. (2015). Children's evaluations of resource allocation in the context of group norms. *Developmental Psychology*, **51**, 554–563. doi:10.1037/a0038796

Eagly, A.H., & Wood, W. (2013). The nature–nurture debates: 25 years of challenges in understanding the psychology of gender. *Perspectives on Psychological Science*, **8**, 340–357. doi:10.1177/1745691613484767

Horn, S.S. (2007). Adolescents' acceptance of same-sex peers based on sexual orientation and gender expression. *Journal of Youth and Adolescence*, **36**, 363–371. doi:10.1007/s10964-006-9111-0

Horn, S.S. (2008). The multifaceted nature of sexual prejudice: how adolescents reason about sexual orientation and sexual

prejudice. In S.R. Levy & M. Killen (Eds.), *Intergroup attitudes and relations in childhood through adulthood* (pp. 173–190). Oxford: Oxford University Press.

Horn, S.S., & Sinno, S. (2014). Gender, sexual orientation and discrimination based on gender and sexual orientation. In M. Killen & J.G. Smetana (Eds.), *The handbook of moral development* (2nd edn.) (pp. 317–339). New York: Taylor and Francis.

Killen, M., Mulvey, K.L., Richardson, C., Jampol, N., & Woodward, A. (2011). The accidental transgressor: Morally-relevant theory of mind. *Cognition*, **119**, 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**, 587–596. doi:10.1037/0012-1649.37.5.587

Killen, M., & Rutland, A. (2011). *Children and social exclusion: Morality, prejudice, and group identity*. New York: Wiley/Blackwell Publishers.

Killen, M., Rutland, A., Abrams, D., Mulvey, K.L., & Hitti, A. (2013). Development of intra- and intergroup judgments in the context of moral and social-conventional norms. *Child Development*, **84**, 1063–1080. doi:10.1111/cdev.12011

Killen, M., & Smetana, J.G. (2015). Origins and development of morality. In M.E. Lamb (Ed.), *Handbook of child psychology and developmental science* (7th edn., Vol. 3, pp. 701–749). New York: Wiley-Blackwell.

Lamb, L.M., Bigler, R.S., Liben, L.S., & Green, V.A. (2009). Teaching children to confront peers' sexist remarks: Implications for theories of gender development and educational practice. *Sex Roles*, **61**, 361–382. doi:10.1007/s11199-009-9634-4

Leman, P.J., Ahmed, S., & Ozarow, L. (2005). Gender, gender relations, and the social dynamics of children's conversations. *Developmental Psychology*, **41**, 64–74.

Liben, L.S., & Bigler, R.S. (2002). The developmental course of gender differentiation: conceptualizing, measuring and evaluating constructs and pathways. *Monographs of the Society for Research in Child Development*, **67**. doi:10.1111/1540-5834.t01-1-00187

Martin, C.L., Eisenbud, L., & Rose, H. (1995). Children's gender-based reasoning about toys. *Child Development*, **66**, 1453–1471. doi:10.1111/j.1467-8624.1995.tb00945.x

Martin, C.L., & Ruble, D. (2004). Children's search for gender cues: cognitive perspectives on gender development. *Current Directions in Psychological Science*, **13**, 67–70. doi:10.1111/j.0963-7214.2004.00276.x

Moses, L.J., Carlson, S.M., & Sabbagh, M.A. (2005). On the specificity of the relation between executive function and children's theories of mind. In W. Schneider, R. Schumann-Hengsteler & B. Sodian (Eds.), *Young children's cognitive development: Interrelationships among executive functioning, working memory, verbal ability, and theory of mind* (pp. 131–145). Mahwah, NJ: Lawrence Erlbaum Associates.

Mulvey, K.L., Hitti, A., & Killen, M. (2013). Intentionality, morality, and exclusion: How children navigate the social world. In M. Banaji & S. Gelman (Eds.), *Navigating the social world: What infants, children and other species can teach us* (pp. 377–384). Oxford: Oxford University Press.

Mulvey, K.L., Hitti, A., Rutland, A., Abrams, D., & Killen, M. (2014). Reasoning about resource allocation in an intergroup context from an individual and a group perspective. *Journal of Social Issues*, **70**, 29–46. doi:10.1111/josi.12045

Nesdale, D., & Lawson, M.J. (2011). Social groups and children's intergroup attitudes: Can school norms moderate the effects of social group norms? *Child Development*, **82**, 1594–1606.

Nesdale, D., Lawson, M.J., Durkin, K., & Duffy, A. (2010). Effects of information about group members on young children's attitudes towards the in-group and out-group. *British Journal of Developmental Psychology*, **28**, 467–482. doi:10.1348/026151009x433321

Posada, R., & Wainryb, C. (2008). Moral development in a violent society: Colombian children's judgments in the context of survival and revenge. *Child Development*, **79**, 882–898. doi:10.1111/j.1467-8624.2008.01165.x

Rochat, P., & Ferreira, C.P. (2008). *Homo negotiatuus: ontogeny of the unique ways humans own, share and reciprocate*. In S. Itakura & K. Fujita (Eds.), *Origins of the social mind: Evolutionary and developmental views* (pp. 141–156). New York: Springer Science + Business Media.

Ruble, D.N., Martin, C.L., & Berenbaum, S. (2006). Gender development. In W. Damon & N. Eisenberg (Eds.), *Handbook of child psychology: Vol. 3, Personality and social development* (6th edn., pp. 141–156). New York: Springer Science + Business Media.

Slaughter, V., Peterson, C.C., & Moore, C. (2013). I can talk you into it: Theory of mind and persuasion behavior in young children. *Developmental Psychology*, **49**, 227–231. doi:10.1037/a0028280

Smetana, J.G. (1986). Preschool children's conceptions of sex-role transgressions. *Child Development*, **57**, 832–871.

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 edn., pp. 23–45). New York: Psychology Press.

Smetana, J.G., Jambon, M., Conry-Murray, C., & Sturge-Apple, M. (2012). Reciprocal associations between young children's moral judgments and their developing theory of mind. *Developmental Psychology*, **48**, 1144–1155. doi:10.1037/a0025891

Takagishi, H., Kameshima, S., Schug, J., Koizumi, M., & Yamagishi, T. (2010). Theory of mind enhances preference for fairness. *Journal of Experimental Child Psychology*, **105**, 130–137. doi:10.1016/j.jecp.2009.09.005

Trautner, H.M., Ruble, D.N., Cyphers, L., Kirsten, B., Behrendt, R. et al. (2005). Rigidity and flexibility of gender stereotypes in childhood: Developmental or differential? *Infant and Child Development*, **14**, 365–381. doi:10.1002/icd.399

Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge: Cambridge University Press.

Watson, A.C., Nixon, C.L., Wilson, A., & Capage, L. (1999). Social interaction skills and theory of mind in young children. *Developmental Psychology, 35*, 386–391. doi:10.1037/0012-1649.35.2.386

Wellman, H.M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development, 75*, 502–517. doi:10.1111/j.1467-8624.2004.00691.x

Zemore, S.E., Fiske, S.T., & Kim, H.-J. (2000). Gender stereotypes and the dynamics of social interaction. In T. Eckes & H.M. Trautner (Eds.), *The developmental social psychology of gender* (pp. 207–241). Mahwah, NJ: Lawrence Erlbaum Associates.

Zucker, K., Wilson-Smith, D., Kurita, J., & Stern, A. (1995). Children's appraisals of sex-typed behavior in their peers. *Sex Roles, 33*, 703–725. doi:10.1007/bf01544775

Received: 20 June 2014

Accepted: 24 June 2015