CHILD DEVELOPMENT

Child Development, January/February 2019, Volume 90, Number 1, Pages 127–135

The Reciprocal Relations Between Female and Male Play Partners and Aggression in Early Childhood

Kristin J. Perry and Jamie M. Ostrov
University at Buffalo and The State University of New York

This study examined the relations between male and female play partners (PP) and relational (RA) and physical (PA) aggression in an early childhood short-term longitudinal study ($N = 164$, $M_{age} = 47.11$ months, $SD = 7.37$ months). A cross-lagged path analysis was used to examine these relations from Time 1 (T1) to Time 2 (T2), 4 months after T1, and a multiple group analysis was tested across gender. Results showed that T1 male PP predicted an increase in PA, T1 RA predicted an increase in female PP, and T1 PA predicted a decrease in female PP. One path was not equivalent across gender. A post hoc interaction between male and female PP at T1 on future RA was evaluated.

In the past decade, researchers have made great strides in understanding gender differences in aggression. However, this research has not examined the intersection of gender development and aggression. Specifically, gender segregated play, which begins to emerge in early childhood and emphasizes that the importance of peer relationships is a critical milestone for children's development (Maccoby, 1990). Early childhood is also an important time for the development of aggression, as physical aggression (PA) is still prevalent and nonphysical forms of aggression are just developing (Côté, Vaillancourt, LeBlanc, Nagin, & Tremblay, 2006; Crick et al., 2006). Relational aggression (RA) is the intent to harm others through the relationship or threat of removal of the relationship (e.g., exclusion; Crick & Grotpeter, 1995), and PA is the intent to harm others through physical acts (Eisner & Malti, 2015). This study examines the bidirectional relations between the gender of a child's play partners (PP) and the type of aggression a child displays.

The gender-linked model of aggression subtypes (i.e., the gender informed social information processing model) is one model that emphasizes the dynamic relation between the child's gender, past peer experiences, and the form of aggression the child exhibits (Ostrov & Godleski, 2010). This theory is predicated on the social information processing model (SIP; Crick & Dodge, 1994), which posits that a child has a database of past social experiences that inform processing and decision making about future social experiences. This is hypothesized to be a cyclical six-step process formed by: (a) encoding of cues, (b) interpretation of cues, (c) clarification of goals, (d) response access or construction, (e) response decision, and (f) behavioral enactment (Crick & Dodge, 1994). The gender-linked model of aggression subtypes theorizes that the SIP model is moderated by gender, which is a socially constructed variable (Ostrov & Godleski, 2010). According to the model, developmentally salient gender schemas influence a child's social knowledge and what cues the child attends to, interprets, and encodes. Moreover, peer interactions contribute to gender schemas where gender-inconsistent acts lead to more negative responses from peers than gender consistent acts, which influence future behavioral enactment (Ostrov & Godleski, 2010). This model has been previously supported (e.g., Ettekal & Ladd, 2015; Spieker et al., 2012).

Within this framework, social schemas are shaped by socialization agents whom model gender appropriate aggressive behavior and contribute to a child's database of past experiences (Crick & Dodge, 1994). In early childhood, PP are important

We thank the UB Social Development Lab for their assistance, and special thanks to Dr. Kimberly Kamper-DeMarco and Sarah Blakely-McClure for methodological assistance and coordination of the data collection. We are grateful to the families, teachers, and administrators of participating schools. The study was funded in part by the Mark Diamond Fund at the University at Buffalo. Preparation of this manuscript was facilitated by a grant from the National Science Foundation to the second author (BCS-1450777).

Correspondence concerning this article should be addressed to Kristin J. Perry, Department of Psychology, University at Buffalo, 478 Park Hall, Buffalo, NY 14260-4110. Electronic mail may be sent to kperry5@buffalo.edu.

 0009-3920/2019/9001-0013
DOI: 10.1111/cdev.13178

socialization agents for gender development and the development of aggressive behavior (Leaper, 2015; Ostrov & Godleski, 2010). Ostrov and Godleski (2010) theorize and researchers have found evidence that RA is the modal or most common form of aggression for young girls (e.g., Ostrov, Kamper, Hart, Godleski, & Blakely-McClure, 2014), and boys exhibit more PA than girls (Card, Stucky, Sawalani, & Little, 2008). Therefore, it is hypothesized that interactions with female peers will predict an increase in relationally aggressive behavior, as female peers will reinforce this behavior, and interactions with male peers will predict an increase in physically aggressive behavior, as male peers will reinforce this behavior.

In Step 5 of the SIP model, the child decides how to act within their current social interactions, which leads to Step 6, behavioral enactment (Crick & Dodge, 1994). A child's gender schema may play a role during these stages, where the child decides whether the behavior is gender consistent. Girls are more likely to endorse RA as gender consistent, and boys are more likely to endorse PA as gender consistent (Giles & Heyman, 2005). Based on the role of these gender schemas, it is hypothesized that pathways from female and male PP to aggression will be moderated by gender. Specifically, male PP will more strongly predict PA for boys and female PP will more strongly predict RA for girls. To statistically examine this effect, the equivalence of regression pathways in the model will be tested across gender.

Finally, children receive feedback from their peer(s) after engaging in an aggressive act (Crick & Dodge, 1994). This feedback gets transcribed into gender and social schemas that influence the child's future social interactions. Researchers have also theorized that there is a gender segregation cycle, where a child interacts with a same-gender peer, learns what behaviors this same-gender peer is exhibiting, and begins to engage in this type of behavior, which leads to further interactions with same-gender peers (Martin, Fabes, & Hanish, 2014). Therefore, children who exhibit RA may receive more positive feedback from female peers, resulting in increased interactions with female peers, and negative feedback from male peers. Similarly children who exhibit PA may receive positive feedback from male peers, resulting in increased future interactions with male peers and negative feedback from female peers. Thus, it is hypothesized that PA will predict a decrease in female PP and an increase in male PP, and RA will predict an increase in female PP but not male PP.

Method

Participants

This study included 164 participants (50.9% boys, M_{age} = 47.75 months, SD = 7.37 months) from two separate independent samples conducted over a 2-year period (Ostrov, Godleski, Kamper-DeMarco, Blakely-McClure, & Celenza, 2015; Perry & Ostrov, 2018). Children were from relatively diverse backgrounds (4.9% African American, 9.9% Asian, Pacific Islander, or Indian, 3.7% Hispanic or Latino, 13.6% multiracial, 67.9% White). Parental occupation was coded using Hollingshead's (1975) four-factor index 9-point scoring system. Parents' education was not taken and was not included in the total factor score. Values ranged from 2 to 9 with a 7.00 average, suggesting that a typical family in our sample was from the third highest occupation group (i.e., 7 = small business owners, minor professionals). Children were recruited from nine National Association for the Education of Young Children accredited or recently accredited early childhood centers in a large northeastern US city (16 classrooms).

The first sample used data from the control group of an intervention study (n = 61, M_{age} = 48.81 months, SD = 7.01, 55.7% girls). Data for the second sample were collected a year after the first sample with children who were not in the intervention sample (n = 105, M_{age} = 47.37 months, SD = 7.56, 47.6% girls). The two samples were completely independent and there were no differences between the two samples for any demographic variables, such as, ethnicity, $\chi^2(6) = 2.73$, $p = .84$, occupation, $\chi^2(6) = 8.13$, $p = .23$, gender, $\chi^2(1) = 0.91$, $p = .34$, or age, $t(149) = 1.15$, $p = .25$.

Procedure

All children in participating classrooms were invited to participate and parents provided written consent for their children. Head teachers also provided written consent prior to report completion. This study was approved by the local institutional review board. Teachers were compensated \$10–\$25 depending on their class size at each time point. Time 1 (T1) data were collected in the fall and Time 2 (T2) data were collected in the spring of the school year, approximately 4 months after T1 data collection was completed.

Measures

Observations of PP

Trained undergraduate and graduate research assistants collected naturalistic observations using a

focal child sampling with continuous recording procedure (Ostrov & Keating, 2004). Prior to classroom entry, consistent with prior procedures (see Ostrov & Keating, 2004), observers underwent extensive training. Observers were trained to identify PP for this study as well as aggression and prosocial behavior for the larger project. Typically there were two to three observers per classroom. Observations were undertaken in a 2-month period, with the goal of completing eight, 10-min observation sessions per child. A peer was considered a PP if there was an interaction with the focal child that lasted at least three seconds (Ostrov & Keating, 2004). Peers could be counted as PP multiple times as long as there was a break in time between interactions. An average number of male and female PP was generated by dividing the total number of female and male PP by the number of completed sessions for the child (i.e., up to eight sessions). On average, each child had a total of 7.59 (SD = 1.03) 10-min observations at the end of T1 and 7.83 (SD = 0.66) observations at the end of T2. Reliability sessions were collected throughout the study for 19.2% of observations at T1 and 18.6% of observations at T2, which is within an acceptable range of interrater reliability sampling percentages (e.g., Ostrov & Hart, 2012; Pellegrini et al., 2011). These sessions demonstrated that observations of male and female PP were reliable at T1 and T2 (all intraclass correlations with absolute agreement > .91). Potential PP in the classroom ranged between 9 and 18 children (M = 14.20, SD = 2.07). This variable was statistically controlled in the primary analyses.

Preschool Social Behavior Scale-Observer and Teacher Report

Physica aggression and RA were measured with the Preschool Social Behavior Scale-Observer Report (PSBS-OR; Ostrov, 2008) and Teacher Report (PSBS-TR; Crick, Casas, & Mosher, 1997). The physical aggression and RA subscales each contained six items rated on a 5-point Likert scale (1 = *never or almost never true*; 5 = *always or almost always true*). After completing behavioral observations, one undergraduate observer from each classroom was randomly selected to complete reports for each participant, which comprise the observer reports. Teacher reports were also completed at the end of the observation period. The undergraduate observers and teachers were not privy to the study hypotheses. Prior research has validated the use of observer reports (e.g., Murray-Close & Ostrov, 2009). The subscales of the PSBS-OR and PSBS-TR had good internal consistency in this sample (Cronbach's α's > .91).

Data Analysis Plan

First, descriptive data of the measures were obtained. Outliers were modified by adjusting the value to ±3 SDs from the mean (Kline, 2011). Skew statistics were assessed, where skew values ranged from 0.32 to 1.46 and kurtosis statistics ranged from −1.17 to 1.50, which suggests that any nonnormality in the data did not impact the analyses (Kline, 2011). Maximum likelihood estimation was used, and missing data were accommodated using full information maximum likelihood. Ten children had missing observer report or observational data, and 15 children had missing teacher report data at T1 or T2. There were no differences in children who had missing data compared to children who had full data for observer or teacher reported RA or PA, male or female PP, or demographic variables. These statistics are available in Supporting Information. However, children with missing teacher data were more likely to be in Study 2 rather than Study 1, $\chi^2(1) = 4.02$, $p = .045$, suggesting that missing teacher report data may not be missing at random. Therefore, study was controlled for in the teacher analyses.

All models were estimated in Mplus version 7.4 (Muthén & Muthén, 1998–2015). A cross-lagged path analysis was used to examine the reciprocal relations between PP and aggression controlling for age. Next, a multiple group analysis across gender was tested. First, a measurement model was tested to examine model fit in the entire sample. Second, a model was tested in which the regression paths were free to vary for boys and girls. Finally, a model was tested, in which the regression paths were constrained to equivalence across gender. Partial equivalence was assessed using sequential use of modification indices (MI) to determine which parameters should be sequentially freed in accordance with procedures outlined by Yoon and Millsap (2007).

A likelihood ratio chi-square test was used to test overall model fit where $p > .05$ indicates good model fit. The following alternative fit indices were also considered: (a) comparative fit index (CFI), where values > .95 suggest good fit, (b) standardized root mean square residual (SRMR) where values < .08 represent mediocre fit, and values < .05 indicate close fit (Hu & Bentler, 1999), and (c) root mean square error of approximation (RMSEA; Steiger, 1990), where values > .10 represent

unacceptable fit, values < .08 suggest mediocre fit, and values < .05 indicate close fit (Browne & Cudeck, 1992; MacCallum, Browne, & Sugawara, 1996). To test comparisons in model fit, chi-square difference tests were used. These models were replicated using teacher reports of aggression. Observers were used as the main informants of aggressive behavior because they are trained to identify aggression, there was less missing data using observer report, and the observer reports were missing at random, whereas teacher reports were not.

Results

Preliminary Analyses

See Table 1 for descriptive statistics and correlations among the variables. All preliminary analyses used observer reports of aggression. First, between group gender differences were examined for T1 variables. Girls had more female PP, $F(1, 163) = 43.59$, $p < .001$, $\eta_p^2 = .21$, fewer male PP, $F(1, 163) = 31.21$, $p < .001$, $\eta_p^2 = .16$, and less PA, $F(1, 162) = 7.94$, $p = .005$, $\eta_p^2 = .05$, at T1 than boys. There was a trend toward girls having higher T1 RA compared to boys, $F(1, 162) = 3.81$, $p = .053$, $\eta_p^2 = .02$. Next between group gender differences were examined for T2 variables. Girls had a greater number of T2 female PP, $F(1, 151) = 20.96$, $p < .001$, $\eta_p^2 = .12$, fewer T2 male PP, $F(1, 151) = 6.99$, $p = .009$, $\eta_p^2 = .04$, higher T2 RA, $F(1, 151) = 9.49$, $p = .002$, $\eta_p^2 = .06$, compared to boys when controlling for the pertinent T1 variable. There was no gender difference in T2 PA, $F(1, 161) = 0.20$, $p = .66$, $\eta_p^2 = .001$. Finally, within group gender differences were examined at T1 using paired samples *t*-tests. Boys interacted with more male PP than female PP at T1, $t(82) = 5.66$, $p < .001$, $d = .58$, and had no difference between relational and PA at T1, $t(82) = -0.80$, $p = .43$, $d = .08$. Girls had more female PP than male PP at T1, $t(81) = -6.35$, $p < .001$, $d = .81$, and had higher RA scores compared to PA scores at T1, $t(80) = 7.28$, $p < .001$, $d = .73$.

Primary Analyses

The measurement model was an adequate fit to the data, $\chi^2(5) = 13.94$, $p = .02$, CFI = .98, SRMR = .05, RMSEA = .10. Results suggested that T1 female PP predicted T2 male PP ($\beta = -.17$, $p < .01$), T1 male PP predicted T2 PA ($\beta = .20$, $p < .01$), and T1 male PP ($\beta = -.18$, $p = .01$), T1 PA ($\beta = -.24$, $p < .01$), and T1 RA ($\beta = .20$, $p = .02$) predicted T2 female PP controlling for age and potential PP. All variables were stable from T1 to T2 (βs range from .42 to .57, $p < .001$). Next, a model was tested in which the regression paths were free to vary across gender, which was an adequate fit to the data, $\chi^2(10) = 18.45$, $p = .05$, CFI = .97, SRMR = .06, RMSEA = .10. Subsequently, a model was tested in which the regression paths were constrained to equivalence across gender. This model was a poor fit to the data, $\chi^2(34) = 58.69$, $p = .01$, CFI = .92, SRMR = .08, RMSEA = .09, and was a significantly worse fit than the free to vary model, $\Delta\chi^2(24) = 40.24$, $p = .02$. To determine whether there was partial nonequivalence across paths, MI were examined until there was no difference in model fit between the new model and the free to vary model or remaining MI values were below 3.84 (Whittaker, 2012). A MI of 3.84 or higher indicates that there will be a significant ($p < .05$) reduction in the chi-squared value if that parameter is freed. In the constrained to equivalence model, the parameter from T1 female PP to T2 RA had the highest MI (MI = 10.56). Freeing this parameter across gender led to an adequate fit to the data, $\chi^2(33) = 47.37$, $p = .05$, CFI = .96, SRMR = .077, RMSEA = .07, and there was no difference in model fit between this model and the free to vary model, $\Delta\chi^2(23) = 28.92$, $p = .18$. For boys, T1 female PP predicted a decrease in T2 RA ($\beta = -.29$, $p < .01$) and for girls, there was a nonsignificant trend toward T1 female PP predicting an increase in T2 RA ($\beta = .14$, $p = .10$). The standardized regression coefficients of the finalized model are shown in Figure 1.

Replication

All models using teacher reported aggression were an adequate fit to the data. The model fit statistics and a figure are shown in Supporting Information. All paths from the prior model were replicated except for the path from T1 male PP to T2 PA. Similar to the original model, one path was moderated by gender such that for girls, T1 PA predicted a decrease in T2 RA ($\beta = -.35$, $p < .01$) and for boys, this relation was not significant ($\beta = .17$, $p = .29$). Additionally, in the teacher model, T1 PA was predictive of T2 male PP ($\beta = .19$, $p = .049$).

Post Hoc Analyses

Consistent with the gender informed SIP model, the impact of female PP on future RA varied by gender. To follow-up on this finding, we examined

Table 1
Descriptive Statistics and Correlations

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.
1. Age (in months)	—												
2. RA T1-observer report	.22**	—											
3. PA T1-observer report	.10	.61**	—										
4. RA T1-teacher report	.01	.27**	.31**	—									
5. PA T1-teacher report	.18*	.23**	.53**	.70**	—								
6. Female PP T1	.13	.11	−.02	.003	.03	—							
7. Male PP T1	.09	.12	.22**	.08	.18*	−.21**	—						
8. RA T2-observer report	.43**	.56**	.28**	.18*	.21**	.15	.04	—					
9. PA T2-observer report	.20*	.43**	.64**	.28**	.50**	.00	.32**	.53**	—				
10. RA T2-teacher report	−.02	.23**	.28**	.56**	.37**	−.05	.04	.22**	.32**	—			
11. PA T2-teacher report	.06	.20*	.49**	.51**	.70**	−.08	.19*	.20*	.57**	.63**	—		
12. Female PP T2	.32**	.16*	−.12	.11	−.01	.50**	−.21**	.32**	.03	.07	−.09	—	
13. Male PP T2	.22**	.06	.14	.17*	.31**	−.24**	.56**	.17*	.33**	.20*	.28**	−.21*	—
M	47.88	1.76	1.55	1.76	1.57	2.50	2.45	1.74	1.59	1.91	1.62	2.64	2.66
SD	7.38	0.77	0.68	0.76	0.81	1.36	1.19	0.73	0.63	0.80	0.76	1.38	1.43
Range	35–63	1–4	1–3.69	1–4.04	1–4.03	0–6.62	.13–6.0	1–3.83	1–3.48	1–3.83	1–3.90	0.20–6.5	0.25–6.96

Note. PP values represent the average number of PP per observation session. PA = physical aggression; RA = relational aggression; PP = play partners; T1 = Time 1; T2 = Time 2.
*$p < .05$. **$p < .01$.

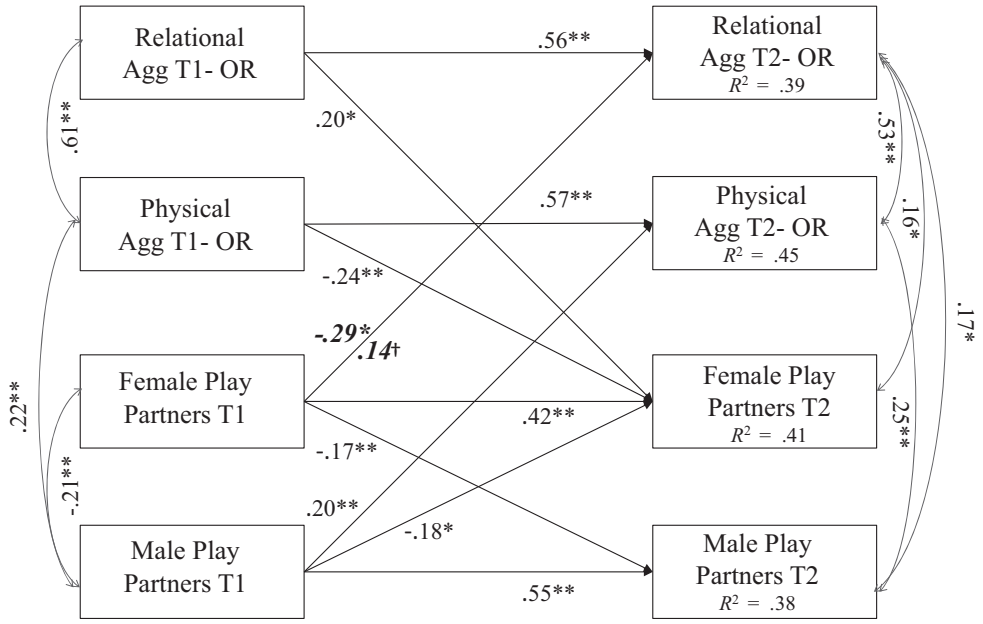

Figure 1. Significant standardized regression coefficients are shown but all within time associations were also tested and included in the model. Bold italicized values represent the gender nonequivalent parameter, where boys are above the diagonal and girls are below the diagonal. R^2 values represent the total amount of variance explained in the T2 variables. Play partner (PP) values represent the average number of PP per observation session. Age and potential PP were statistically controlled. T1 = Time 1; T2 = Time 2; Agg = aggression; OR = observer report.
$^{\dagger}p < .10$. $*p < .05$. $**p < .01$.

whether there was an interaction between male and female PP on future RA. According to the aforementioned model, it would be important for boys exposed to high levels of female peers to also be exposed to high levels of male peers to truly learn the contrast between male and female peer group behaviors. This interaction was tested controlling for all regression paths in the finalized model. The simple slopes for the interaction were tested using Preacher and colleagues' method (Preacher, Curran, & Bauer, 2006) designed to identify the region of significance of the interaction or at what points of the moderator (i.e., male PP at T1), the relation between female PP at T1 (IV) and RA at T2 (DV) is significant. The upper bound represents the positive point of male PP (moderator) that the female PP T1 (IV) and RA T2 (DV) association becomes significant, and the lower bound represents the lowest point at which this relation becomes significant. All variables were centered.

The model with the new interaction coefficient was an adequate fit to the data, $\chi^2(17) = 34.10$, $p = .01$, CFI = .96, SRMR = .04, RMSEA = .08. The interaction was significant ($\beta = -.16$, $p < .01$) controlling for all other variables. The lower bound of the region of significance was −0.47 male peers ($B = .15$, $SE = .08$, $p = .05$), which is an impossible value in our data as the fewest number of male PP a child could have is zero. The upper bound of the region of significance was 2.83 ($B = -.07$, $SE = .04$, $p = .05$), which suggests that when there are 2.83 male PP or more per observation session, there is a negative relation between T1 female PP (IV) and T2 RA (DV). This negative association gets stronger as the number of male PP (moderator) increases. Simple slopes for the minimum (0.13 male PP) and the maximum (6 male PP) were probed and are shown in Figure 2. This interaction was not significant when using teacher reports of aggression ($\beta = .06$, $p = .30$).

A three-way interaction between T1 male PP, T1 female PP, and gender on T2 observer reported RA was examined. All lower order interaction terms and paths from the measurement model were included. The model was an adequate fit to the data, $\chi^2(25) = 43.11$, $p = .01$, CFI = .96, SRMR = .04, RMSEA = .07. The three-way interaction was not significant ($\beta = -.07$, $p = .28$).

The models were also run accounting for clustering of children within classrooms ($n = 16$). The same paths and nonequivalent paths emerged as significant in all models when accounting for clustering, with the exception of one parameter in the observer model [i.e., T1 RA to T2 female PP became

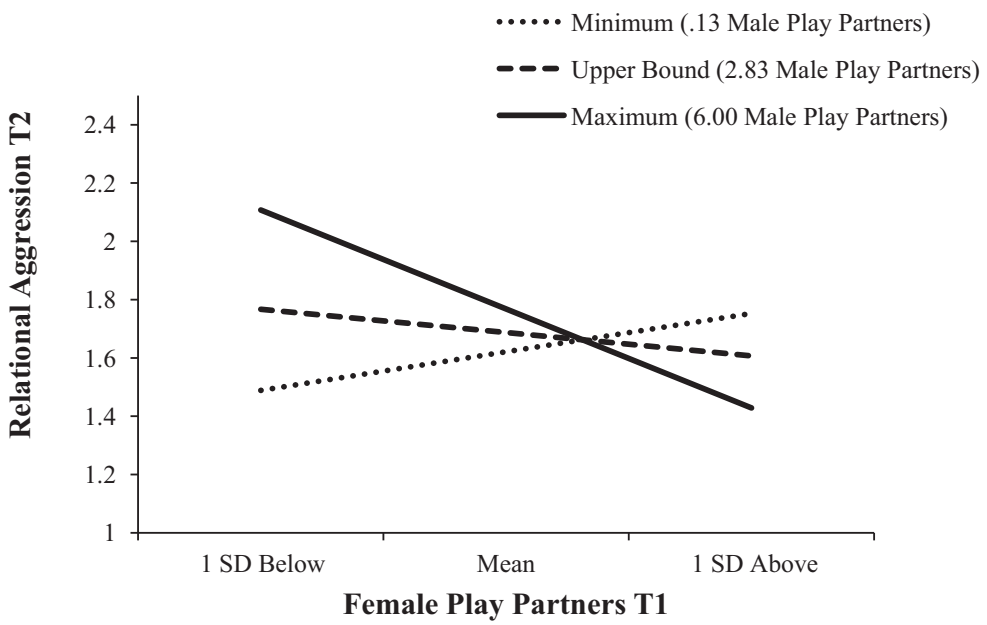

Figure 2. This figure illustrates the interaction between female play partners (PP; IV) and male PP (moderator) at Time 1 (T1) on Time 2 (T2) relational aggression (DV). Relational aggression at T1 was controlled. At the maximum value of male PP (moderator), the IV–DV association is significant ($B = -.28$, $SE = .09$, $p = .002$) and at the minimum value of male PP (moderator), the IV to DV association is a nonsignificant trend ($B = .11$, $SE = .07$, $p = .086$).

nonsignificant ($B = .35$, $SE = .19$, $p = .065$)]. However, results from the models that do not account for clustering are reported because at least 20 clusters are needed to accurately run a cluster analysis (Muthén, 2018) and the standard errors of the parameters are potentially untrustworthy when running the current model with 16 clusters.

Discussion

This study examined the reciprocal relations between the gender of a child's PP and RA and PA. Preliminary results demonstrated that girls experience a greater increase in female PP relative to boys, and boys experience a greater increase in male PP relative to girls, providing support for gender segregation processes beginning to occur in preschool. Primary results found that PA predicted a decrease in female PP, RA predicted an increase in female PP, and male PP predicted an increase in future PA. These results were replicated when using teacher reports of aggressive behavior with the exception of the path from male PP to future PA. These findings are consistent with the gender informed SIP model, such that girls tend to exhibit more RA relative to PA, and therefore in female peer groups PA may be an undesirable behavior and RA may be perpetuated through modeling (Bandura, 1973; Maccoby, 1998; Snethen & Van Puymbroeck, 2008).

One path was moderated by gender, such that the effect of T1 female peers on T2 RA varied for boys and girls. Female PP predicted a decrease in RA for boys and an increase in RA for girls (this was a trend). To understand this moderation further we examined whether there was an interaction between male and female PP at T1 on future RA. Results showed that male PP moderated the effect of female PP on future RA. Consistent with the gender informed SIP model (Leaper, 2015; Ostrov & Godleski, 2010), boys who have a high level of male and female PP may experience more situations in which to gather information about what is gender consistent behavior and may come to associate PA as a "boy behavior" and RA as a "girl behavior." Overall, these results indicate that gender schemas are likely at play in the relation between female PP and future RA. This was not found for the relation between male PP and PA suggesting that there is something unique about female peers and RA. Prior research has found that girls have a stronger peer preference for same-gendered peers than boys (Gasparini, Sette, Baumgartner, Martin, & Fabes, 2015) and there are more consequences for boys breaking stereotypes than girls (Mulvey & Killen, 2015), indicating that girls may be driving gender-differentiated aggressive behaviors. However, these

moderation results should be interpreted with caution, as they were not replicated when using teacher reports.

Despite the novelty and intensity of the observational method used there are limitations to this study. The participants were from a limited number of schools in the northeastern United States and results may not be generalizable. Similarly, there was a short interval between time points, and we were unable to study whether the impact of gender segregation translates to new peer groups over time. In regards to methodology, with a small number of classrooms, nesting effects were not evaluated and the interrater reliability sampling rate may have been lower than some recommendations. Additionally, the underlying processes of these effects are unknown (e.g., is it attention to same-gender peers that matters or engagement with peers that matters). Finally, we did not examine the proportion of male to female PP, which is another way to evaluate gender segregation processes. In sum, we found evidence for the gender informed SIP model. Over and above modeling of peers, there may be unique gender-based social-cognitive processes involved in the development of RA. This can inform intervention work by determining how peer based interventions may be most effective in reducing aggression subtypes.

References

Bandura, A. (1973). *Aggression: A social learning analysis*. Englewood Cliffs, NJ: Prentice Hall.

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods and Research*, *21*, 230–258. https://doi.org/10.1177/0049124192021002005

Card, N. A., Stucky, B. D., Sawalani, G. M., & Little, T. D. (2008). Direct and indirect aggression during childhood and adolescence: A meta-analytic review of gender differences, intercorrelations, and relations to maladjustment. *Child Development*, *79*, 1185–1229. https://doi.org/10.1111/j.1467-8624.2008.01184.x

Côté, S. M., Vaillancourt, T., LeBlanc, J. C., Nagin, D. S., & Tremblay, R. E. (2006). The development of physical aggression from toddlerhood to pre-adolescence: A nationwide longitudinal study of Canadian children. *Journal of Abnormal Child Psychology*, *34*, 85. https://doi.org/10.1007/s10802-005-9001-z

Crick, N. R., Casas, J. F., & Mosher, M. (1997). Relational and overt aggression in preschool. *Developmental Psychology*, *33*, 579–588. https://doi.org/10.1037/00121649.33.4.579

Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social information-processing mechanisms in children's social adjustment. *Psychological Bulletin*, *115*, 74–101. https://doi.org/10.1037/0033-2909.115.1.74

Crick, N. R., & Grotpeter, J. K. (1995). Relational aggression, gender, and social psychological adjustment. *Child Development*, *66*, 710–722. https://doi.org/10.2307/1131945

Crick, N. R., Ostrov, J. M., Burr, J. E., Cullerton-Sen, C., Jansen-Yeh, E., & Ralston, P. (2006). A longitudinal study of relational and physical aggression in preschool. *Journal of Applied Developmental Psychology*, *27*, 254–268. https://doi.org/10.1016/j.appdev.2006.02.006

Eisner, M. P., & Malti, T. (2015). Aggressive and violent behavior. In M. E. Lamb (Vol. Ed.) and R.M. Lerner (Series Ed.), *Handbook of child psychology and developmental science, Vol. 3: Social, emotional and personality development* (7th ed., pp. 795–884). New York: Wiley.

Ettekal, I., & Ladd, G. W. (2015). Costs and benefits of children's physical and relational aggression trajectories on peer rejection, acceptance, and friendships: Variations by aggression subtypes, gender, and age. *Developmental Psychology*, *51*, 1756–1770. https://doi.org/10.1037/dev0000057

Gasparini, C., Sette, S., Baumgartner, E., Martin, C. L., & Fabes, R. A. (2015). Gender-biased attitudes and attributions among young Italian children: Relation to peer dyadic interaction. *Sex Roles*, *73*, 427–441. https://doi.org/10.1007/s11199-015-0526-5

Giles, J. W., & Heyman, G. D. (2005). Young children's beliefs about the relationship between gender and aggressive behavior. *Child Development*, *76*, 107–121. https://doi.org/10.1111/j.1467-8624.2005.00833.x

Hollingshead, A. A. (1975). *Four-factor index of social status*. Unpublished manuscript, Yale University.

Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–55. https://doi.org/10.1080/10705519909540118

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: Guilford.

Leaper, C. (2015). Gender and social-cognitive development. In L. S. Liben, U. Müller, R. M. Lerner, L. S. Liben, U. Müller, & R. M. Lerner (Eds.), *Handbook of child psychology and developmental science: Cognitive processes* (pp. 806–853). Hoboken, NJ: Wiley. https://doi.org/10.1002/9781118963418.childpsy219

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*, 130–149. https://doi.org/10.1037/1082-989X.1.2.130

Maccoby, E. E. (1990). Gender and relationships: A developmental account. *American Psychologist*, *45*, 513–520. https://doi.org/10.1037/0003-066X.45.4.513

Maccoby, E. E. (1998). *The two sexes: Growing up apart, coming together*. Cambridge, MA: Belknap Press/Harvard University Press.

Martin, C. L., Fabes, R. A., & Hanish, L. D. (2014). Gendered-peer relationships in educational contexts. In

L. S. Liben & R. S. Bigler (Eds.), *Advances in child development and behavior, Vol. 47: The role of gender in educational contexts and outcomes* (pp. 151–187). San Diego, CA: Elsevier Academic Press. https://doi.org/10.1016/bs.acdb.2014.04.002

Mulvey, K. L., & Killen, M. (2015). Challenging gender stereotypes: Resistance and exclusion. *Child Development*, *86*, 681–694. https://doi.org/10.1111/cdev.12317

Murray-Close, D., & Ostrov, J. M. (2009). A longitudinal study of forms and functions of aggressive behavior in early childhood. *Child Development*, *80*, 828–842. https://doi.org/10.1111/j.1467-8624.2009.01300.x

Muthén, B. O. (2018). *RE: Example for type=complex*. Retrieved from http://www.statmodel.com/discussion/messages/12/776.html?1523500311.

Muthén, L. K., & Muthén, B. O. (1998–2015). *Mplus (version 7.4)*. Los Angeles, CA: Author.

Ostrov, J. M. (2008). Forms of aggression and peer victimization during early childhood: A short-term longitudinal study. *Journal of Abnormal Child Psychology*, *36*, 311–322. https://doi.org/10.1007/s10802-007

Ostrov, J. M., & Godleski, S. A. (2010). Toward an integrated gender-linked model of aggression subtypes in early and middle childhood. *Psychological Review*, *117*, 233–242. https://doi.org/10.1037/a0018070

Ostrov, J. M., Godleski, S. A., Kamper-DeMarco, K. E., Blakely-McClure, S. J., & Celenza, L. (2015). Replication and extension of the early childhood friendship project: Effects on physical and relational bullying. *School Psychology Review*, *44*, 445–463. https://doi.org/10.17105/spr-15-0048.1

Ostrov, J. M., & Hart, E. J. (2012). Observational methods. In T. D. Little (Ed.), *Oxford handbook of quantitative methods* (Vol. 1, pp. 286–304). New York, NY: Oxford University Press.

Ostrov, J. M., Kamper, K. E., Hart, E. J., Godleski, S. A., & Blakely-McClure, S. J. (2014). A gender-balanced approach to the study of peer victimization and aggression subtypes in early childhood. *Development and Psychopathology*, *26*, 575–587. https://doi.org/10.1017/S0954579414000248

Ostrov, J. M., & Keating, C. F. (2004). Gender differences in preschool aggression during free play and structured interactions: An observational study. *Social Development*, *13*, 255–277. https://doi.org/10.1111/j.14679507.2004.000266.x

Pellegrini, A. D., Bohn-Gettler, C. M., Dupuis, D., Hickey, M., Roseth, C., & Solberg, D. (2011). An empirical examination of sex differences in scoring preschool children's aggression. *Journal of Experimental Child Psychology*, *109*, 232–238. https://doi.org/10.1016/j.jecp.2010.11.003

Perry, K. J., & Ostrov, J. M. (2018). Testing a bifactor model of relational and physical aggression in early childhood. *Journal of Psychopathology and Behavioral Assessment*, *40*, 93–106. https://doi.org/10.1007/s10862-017-9623-9

Preacher, K. J., Curran, P. J., & Bauer, D. J. (2006). Computational tools for probing interaction effects in multiple linear regression, multilevel modeling, and latent curve analysis. *Journal of Educational and Behavioral Statistics*, *31*, 437–448. https://doi.org/10.3102/10769986031004437

Snethen, G., & Van Puymbroeck, M. (2008). Girls and physical aggression: Causes, trends, and intervention guided by social learning theory. *Aggression and Violent Behavior*, *13*, 354. https://doi.org/10.1016/j.avb.2008.05.003

Spieker, S. J., Campbell, S. B., Vandergrift, N., Pierce, K. M., Cauffman, E., Susman, E. J., . . . & NICHD ECCRN. (2012). Relational aggression in middle childhood: Predictors and adolescent outcomes. *Social Development*, *21*, 354–375. https://doi.org/10.1111=j.1467-9507.2011.00631

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, *25*, 173–180. https://doi.org/10.1207/s15327906mbr2502_4

Whittaker, T. A. (2012). Using the modification index and standardized expected parameter change for model modification. *Journal of Experimental Education*, *80*, 26–44. https://doi.org/10.1080/00220973.2010.531299

Yoon, M., & Millsap, R. E. (2007). Detecting violations of factorial invariance using data-based specification searches: A Monte Carlo study. *Structural Equation Modeling*, *14*, 435–463. https://doi.org/10.1080/10705510701301677

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's website:

Figure S1. Teacher Report of Aggression Path Analysis Model

Table S1. Missing Data Statistics

Table S2. Teacher Report of Aggression Model Fit Statistics