INTERACTIVE LINKS BETWEEN RELATIONAL AGGRESSION, THEORY OF MIND, AND MORAL DISENGAGEMENT AMONG EARLY ADOLESCENTS

CONSTANTINOS M. KOKKINOS, IOANNA VOULGARIDOU, MARIANNA MANDRALI, AND CHRYSOULA PAROUSIDOU

Democritus University of Thrace

The aim of this study was to investigate possible interactive links between theory of mind (ToM), moral disengagement and relational aggression, using a moderated mediation analysis, with gender as a moderator, in a sample of 120 Greek preadolescents. Results indicated that relational aggression was significantly positively associated with moral disengagement and negatively with ToM. Moderated mediation analyses indicated that boys with deficient ToM were more likely to morally disengage from their actions which in turn resulted in relational aggression, while poor ToM was directly linked to relational aggression only for preadolescent girls. Moral disengagement had a direct effect on relational aggression only for boys, while ToM was found to partially mediate the relationship between moral disengagement and relational aggression only for girls. The results emphasize that the co-morbid effects of socio-cognitive factors should be taken into consideration when relational aggression is explored. © 2016 Wiley Periodicals, Inc.

The quest to better understand aggression as well as the individual characteristics of those involved has evoked significant interest to researchers. Over the last decades, a growing number of studies have provided explanations which have integrated findings from research on emotional, moral, and social cognitive development with relation to aggression (Barchia & Bussay, 2011; Gini, 2006; Gini, Pozzoli, & Hauser, 2011; Mathieson et al., 2011; Voulgaridou & Kokkinos, 2015). Fairly recently, a social cognitive perspective has been used as a framework to elucidate relational aggression (RA; e.g., Crain, Finch, & Foster, 2005; Mathieson et al., 2011), an aggressive form referring to harming others through the manipulation of peer relationships, including acts such as rumor spreading, damaging someone’s social status, or excluding from a social group (Orpinas, McNicholas, & Nahapetyan, 2014).

To date, studies relative to this issue are limited in two ways. First, there is partial consensus on the actual role of such factors. A debate exists in the literature as to whether aggressive behaviors (especially relational and proactive forms of aggression) constitute a deviant process of social information (Crick & Dodge, 1994; Crick, Grotpector, & Bigbee, 2002) or the result of skilled manipulators with sophisticated perspective-taking abilities (Sutton, Smith, & Swettenham, 1999). Second, although increasing evidence confirms the importance of sociocognitive factors in both general and RA, most research examining these associations has focused on the model of social information processing (e.g., Crain et al., 2005; Mathieson et al., 2011). Crain et al. (2005), however, noted that the investigation of the possible relevance of other sociocognitive factors (i.e., moral disengagement [MD]) for the understanding of RA is essential. In accordance with these perspectives, the current study aims to address these limitations and extend previous findings by integrating two sociocognitive variables, which have traditionally been analyzed separately (i.e., theory of mind [ToM], MD) and by exploring whether they act in an interactive manner to reinforce engagement in RA.

ToM refers to the ability of individuals to attribute mental states to themselves and others to explain and predict behavior (Premack & Woodruff, 1978). Increasing evidence supports its role in the explanation of aggressive behavior (e.g., Caravita, Di Blasio, & Salmivalli, 2010; Shakoor et al., 2012; Sutton et al., 1999). Despite the importance researchers place on the development of ToM

Correspondence to: Constantinos M. Kokkinos, Department of Primary Education, School of Education Sciences, Democritus University of Thrace, N. Hili, GR 68100, Alexandroupolis, Greece. E-mail: kkokkino@eled.duth.gr

253
in the manifestation of relationally aggressive behavior (e.g., Sutton et al., 1999), it is surprising that its role in the emergence of RA has not been explicitly addressed. Indirectly, evidence suggests that cognitive sophistication and perspective-taking skills are required to target victims’ social relationships and status (Givens, 2010; Sutton et al., 1999; Xie, Swift, Cairns, & Cairns, 2002). Xie et al. (2002), for example, suggested that perpetrators of RA are capable of balancing their aggressive behaviors with prosocial and cooperative ones in a successful and maneuverable way. Indeed, Gini (2006) reported a positive correlation between all forms of aggression, including relational, and ToM among preadolescents, suggesting that relationally aggressive children possess higher ToM skills. ToM development allows children to be aware of others’ relationships and, potentially, to determine how to harm others and to inflict distress through gossiping, manipulation, or social exclusion (Gomez-Garibello & Talwar, 2015). If certain types of aggressors are identified to use high levels of social cognition, this would have noteworthy implications for intervention strategies. For instance, relational aggressors being cold, manipulative masters of a social situation may be resistant to traditional aggression interventions, including improvement of their social skills. Conversely, they may require new and innovative approaches, namely teaching them how to be prosocial as well as targeting their cognitions and positive alternatives to negative thoughts (Givens, 2010).

ToM has been characterized as a neutral social tool that could be implicated in either aggressive or prosocial behavior. Thus, focusing on the content of mentalizing ability, research should place attention on other individual factors, such as, empathy, morality, and personality, that determine how an enhanced ToM is used (Givens, 2010). Given its significant association with aggressive behavior (Gini, Pozzoli, & Hymel, 2014), as well as the emerging evidence supporting its role in RA (Kipritsi & Kokkinos, 2015; Kokkinos, Voulgaridou, & Markos, 2014; Murray-Close, Crick, & Galotti, 2006), morality cannot be excluded from this investigation. Social cognitive theory posits that people adopt moral standards that serve as a guide to their behavior. When their actions are against these standards, their self-worth is damaged. To avoid this unpleasant feeling, they develop social and psychological maneuvers so that they can become easily disengaged from their moral standards (Bandura, 1999). According to Bandura (2002), MD is a process that allows the moralization of actions that would otherwise be regarded as immoral through a series of cognitive self-regulatory mechanisms (Gini et al., 2011). As Obermann (2011) pointed out, aggressors were more likely to report higher levels of MD than were their nonaggressive peers. To date, little work has examined how these moral mechanisms are implicated in relationally aggressive conduct (Kipritsi & Kokkinos, 2015; Kokkinos et al., 2014; Murray-Close et al., 2006), taking into consideration that during the late elementary school years, children rate RA as less erroneous and harmful than physical, as they cannot interpret the extent of damage it may cause (Gini, 2006).

Hence, the present study examined the interplay between ToM and MD in explaining RA and accordingly was purported to provide an integrated model of RA in preadolescents. Based on recent evidence, not only do ToM skills precede MD, but moral judgments also actually influence ToM perceptions (e.g., Adams & Steadman, 2004; Knobe, 2005). In an attempt to provide insight into the understanding of the way ToM and MD are coordinated to affect relationally aggressive behavior, the current study explored two different mediation pathway models. In the first, MD serves as a mediator between ToM and RA, whereas, in the second, MD is posited to affect RA both directly and indirectly through its effects on ToM. Gender is found to play a crucial role in the association between MD and ToM on the one side, and RA on the other (e.g., Bosacki, 2013; Gini et al., 2014). In this regard, the study additionally explored the moderating effects of gender.

The present research is of particular importance for school practitioners, given that recent evidence suggests that relational forms of aggression are of major concern among students (McEachern & Snyder, 2012). As Rigby (2004) claims, different perspectives on the causes of aggression have different implications for intervention strategies. To provide a specific example, theorizing RA as the
outcome of individual differences is more likely to direct the attention of educational interventions to disciplinary methods, behavior modification, and counseling. Both educators and school psychologists’ awareness of such links in RA is essential to take proactive steps to decrease this type of behavior in schools (Kerig, 2007) through a sociocognitive intervention pattern. Specifically, such programs would help students explore and practice the identification of feelings, signs of physiological arousal, calming strategies, interpreting intentions of others, and generating and evaluating alternative behaviors to enact (Leff, Waasdorp, & Crick, 2010).

The present study was conducted with early adolescents (10- to 12-year-olds), an age group that has been barely investigated, although assumed at risk for participation in RA (Zimmer-Gembeck, Trevaskis, Nesdale, & Downey, 2014). As preadolescents progress in their sociocognitive abilities (e.g., perspective taking, emotion regulation) they may become increasingly able to use more sophisticated or covert aggressive behaviors (Prinstein, Boergers, & Vernberg, 2001). Besides, research suggests that relationally aggressive behavior, as a more sophisticated way of manipulation, requires an understanding of others’ mental states (Card, Stucky, Sawalani, & Little, 2008), that highlights the necessity of superior ToM skills. Finally, Caravita, Sijtsema, Rambaran, and Gini (2014) have shown that the changes in MD become more prevalent during preadolescence than in late childhood.

**MD MEDIATES THE RELATIONSHIP BETWEEN ToM AND RA**

According to recent integrated models, individual differences in aggression could be attributed to children’s social cognition about aggressive behavior (Gini et al., 2014; Murray-Close et al., 2006), including theories explaining children’s moral reasoning. With regard to ToM, Arsenio and Lemerise (2001) claimed that ToM explanations per se are not likely to be very helpful for understanding the nature of aggression, in that having a sophisticated ToM can lead to either prosocial or aggressive behaviors. In other words, having a superior ToM means nothing about how that knowledge will be used in social interactions (Gini, Abiero, Benelli, & Altoe, 2008). Furthermore, in Gini’s (2006) research, aggressors were found to possess good ToM skills, but they were deficient in moral cognition. However, aggressors’ social skills have also been questioned in studies indicating that poor ToM is linked to higher participation in aggression (Caravita et al., 2010; Shakoor et al., 2012).

The extant literature on aggression has mainly explored direct associations between individual factors and involvement in aggression. However, it has been suggested that individual characteristics, such as cognitions (i.e., ToM, MD), may interact with each other in explaining social behavior and adjustment (Caravita et al., 2010). Following this framework, the first model (Model 1, Figure 1) describes a partial mediation pathway, where MD is posited to partially mediate the effect of ToM on RA.

![Figure 1](image-url)  
**Figure 1.** Hypothesized moderated partial meditational model of theory of mind and moral disengagement as predictors, moral disengagement as partial mediator, relational aggression as outcome, and gender as moderator (Model 1).
Drawing from research on social cognition, it is proposed that relationally aggressive children and adolescents can successfully and deliberately balance their aggressive behaviors with prosocial and cooperative ones (Andreou, 2006). They seem to use subtle skills of manipulation to their advantage for a social attack to be effective (Espelage, Mebane, & Swearer, 2004; Xie et al., 2002). Recently, there has been an expressed interest in whether ToM competence is related to understanding another’s intentions regarding morally relevant actions (Killen, Mulvey, Richardson, Jampol, & Woodward, 2011; Knobe, 2005). Additionally, Baird and Astington (2004) suggested that there is a positive correlation between children’s performance in moral reasoning and false-belief tasks. Prior research has shown that youth who engage in aggressive behavior show deficiencies in moral emotions and have high levels of MD (e.g., Barchia & Bussey, 2011; Caravita et al., 2014). With regard to RA, some evidence shows that MD predicts this aggressive form among preadolescents (Kokkinos et al., 2014). However, as it is stressed by Gini et al. (2014), the association between MD and aggression remains significant even after other predictors of aggressive behavior are taken into consideration. In this regard, it would be theoretically relevant to examine whether ToM predicts not only directly but also indirectly through MD, relationally aggressive behavior (Figure 1).

Practically speaking, intervention strategies emphasizing solely the understanding of others may not be particularly effective, as it is not clear how ToM affects involvement in RA (Givens, 2010). The potential pathway from ToM to RA through MD highlights an additional factor that determines the role of ToM in relationally aggressive behavior (Repacholi, Slaughter, Pritchard, & Gibbs, 2003). Therefore, relational aggressors’ proneness to MD should not be ignored by school psychologists when they apply sociocognitive intervention strategies.

ToM mediates the relationship between MD and RA

In an attempt to better understand the association between MD and RA, the current study additionally explores the mediating role of ToM. It is widely recognized that individuals sometimes use ToM judgments in moral cognition. However, some recent studies (e.g., Knobe, 2005; Leslie, Knobe, & Cohen, 2006), as well as neuroscientific research evidence, has confirmed new connections between ToM and moral judgments, indicating that moral judgments affect intentionality judgments (Killen et al., 2011). This approach deviates from the traditional direction, which perceives ToM skills as prerequisite for moral judgments. Thus, sometimes moral judgments may be applied in the formation of ToM attributions (Knobe, 2005). Leslie et al. (2006) found that children’s interpretations of others’ intentions may follow their moral attributions. It seems that individuals first reach a morally good or bad judgment, which in turn guides them to decisions about the mind (Knobe, 2005). Therefore, it is assumed that children who score high on MD may accordingly exhibit poor ToM skills, as they are more likely to make biased hostile attributions (Model 2, Figure 2). In turn, low ToM is linked to greater engagement in aggression (Shakoor et al., 2012). According to

![Diagram](image-url)

**Figure 2.** Hypothesized moderated partial meditational model of theory of mind and moral disengagement as predictors, theory of mind as partial mediator, relational aggression as outcome, and gender as moderator (Model 2).
Leslie et al. (2006), the influences of morally relevant information on ToM knowledge constitutes a complex mechanism.

The current study examined an additional model of RA in which children’s morality of their behaviors preceded the output judgment regarding others’ intentionality of behaviors. If this pathway is confirmed, it would be of great significance for practitioners, as intervention programs should primarily target perpetrators’ MD. For instance, educators could make efforts to address these distortions in morality, to favor children’s appreciation of the harm they cause, as well as to focus on mitigating the effect of self-justifying mechanisms (Gini et al., 2011; Pozzoli, Gini, & Vieno, 2012).

The Moderating Effects of Gender

To find whether gender increases the strength of the relationship between ToM and RA on the one hand, and MD and RA on the other, participants’ sex was used as a moderator in both models (Figures 1 and 2). Gender has been linked to preadolescents’ RA, resulting in inconsistent evidence (Voulgaridou & Kokkinos, 2015). Some researchers suggest that indirect forms of aggression are more common among girls (Archer & Coyne, 2005), whereas others report no significant differences between boys and girls (Card et al., 2008) or even higher scores in boys (e.g., Juliano, Stetson Werner, & Wright Cassidy, 2006; Kokkinos et al., 2014). Regarding ToM, preadolescent girls perform higher in ToM tasks than do boys, who are more likely to lack in social competence and cognition (Bosacki & Wilde Astington, 1999). However, recent research found no differences between preadolescent boys and girls on advance ToM tasks (i.e., strange stories and faux pas; Kokkinos, Kakarani, & Kolovou, 2015).

Regarding the first link, because RA is associated with social cognition (Kaukiainen et al., 1999), it has been argued that possessing good ToM skills may be linked to aggression, especially among girls (Espelage et al., 2004). In particular, Andreou’s (2006) study showed that RA was associated with the understanding of others’ mental states, as well as with the ability to construe social interactions among early adolescent girls, whereas these associations were weaker or nonexistent among boys of the same age.

In the case of the second association, it was unclear what to anticipate because several studies that have tried to disclose possible gender differences in the association between MD and RA resulted in controversial findings. Thus, there are studies that have revealed no statistically significant differences between boys and girls in this area (Kokkinos et al., 2014), whereas Murray-Close et al. (2006) found that girls were more likely than boys were to consider RA as harmful for the victim, and Paciello, Fida, Tramontano, Lupinetti, and Caprara (2008) found that aggressive boys reported elevated MD. Still, there are some findings suggesting that higher MD may contribute to females’ involvement in RA, but not to males’ (Yadava, Sharma, & Gandhi, 2001).

As far as the association between ToM and MD is concerned, the studies investigating this link (Killen et al., 2011; Knobe, 2005) failed to indicate significant gender differences. In this regard, the moderating effects of gender on the association between ToM and MD will not be explored.

The Present Study

Given the limited number of studies examining the relationship between ToM, MD, and RA during preadolescence, both in Greek and international research literature, the present study aimed at providing evidence regarding the links between the three variables, to better understand the potentially complex mechanisms that underlie early adolescents’ engagement in RA. Specifically, the primary aim of the study was to investigate the association among ToM, MD, and self-reported RA. Considering the existing evidence, it was hypothesized that RA would be positively associated with ToM skills (Hypothesis 1 [H1]; Gini, 2006; Gomez-Garibello & Talwar, 2015; Xie et al., 2002) and MD (Hypothesis 2 [H2]; e.g., Kokkinos et al., 2014; Murray-Close et al., 2006).
However, the current investigation goes beyond testing simple associations by trying to identify the mechanisms governing the interactions between sociocognitive factors and RA. In each case, cognitive variables (MD, ToM) were posited to confer risk for the engagement in RA by acting interchangeably as mediating and predicting variables. As far as the links between the two cognitive factors are concerned, recent studies have found that not only ToM is necessary for moral judgment but also children’s moral decisions regarding attributions of blame influence their interpretations of other’s intentions (Killen et al., 2011). Thus, the purpose of the present study was to extend existing research by examining two different mediating models.

In the first model, ToM was posited to affect RA directly, as well as indirectly through its effects on MD. To make a specific mediational hypothesis, high MD was expected to mediate the association between advanced ToM and RA (Hypothesis 3 [H3]; Baird & Astington, 2004; Killen et al., 2011; Knobe, 2005). However, considering the converging evidence supporting the opposite direction of the association between MD and ToM (Killen et al., 2011), in the second model, MD served as the predictor of RA both directly and via ToM. Therefore, it was hypothesized that high MD will lead to low ToM (Knobe, 2005; Leslie et al., 2006), which in turn will affect relationally aggressive behavior (Caravita et al., 2010; Gini, 2006; Xie et al., 2002; Hypothesis 4 [H4]).

Finally, the study extends prior research by examining whether the aforementioned pathways would function differently for boys and girls, by testing two moderated mediation models. Given the evidence suggesting that the association between ToM and RA prevails for females (Andreou, 2006; Espelage et al., 2004; Kaukiainen et al., 1999), it was expected that high ToM would more strongly predict RA in girls compared to boys (Hypothesis 5 [H5]). Regarding the link between MD and RA, no a priori hypotheses were formulated, given the mixed past evidence. Finally, it is not clear whether the association between MD and RA differs across boys and girls. Although some researchers have found this association stronger among girls (Yadava et al., 2001), others supported an opposite view (Paciello et al., 2008), yet still some others have claimed no moderating gender effects (e.g., Kokkinos et al., 2014; Obermann, 2011).

**METHOD**

**Participants**

A convenience sample of 140 Greek public elementary school students between 10 and 12 years of age, attending fifth and sixth grades, participated in the study. Data were collected only from the region of Thrace, North Eastern Greece, including both public urban and rural schools. All students were native Greeks, and most of them were from middle-class families. Twenty participants were excluded from the sample because they failed to pass the first order false-belief task. Thus, 65 (54.2%) and 55 (45.8%) fifth and sixth graders, almost equally distributed in terms of gender (69 girls; 57.5%) were used in the analyses.

**Materials**

The English versions of MD and RA scales, as well as the faux pas and strange stories tasks, were adapted and translated into Greek according to cross-cultural research guidelines (Beaton, Bombardier, Guillemin, & Ferraz, 2000), using the front and back translation method by employing two bilingual psychologists. For the ToM scale tasks, Mavropoulou’s (personal communication, May 18, 2013) translation was used. The original scoring system for the translated versions was maintained. All the measures were pilot tested in a sample of preadolescents to assess item comprehensibility and duration of administration.

Confirmatory factor analysis (CFA) was used to assess the construct validity of RA and MD scales. In all analyses, the maximum likelihood estimation method and the Satorra-Bentler scaled
chi-square test for non-normal data were used (Mplus version 6.1; Muthén & Muthén, 2010). The model fit was evaluated by means of a number of fit indexes such as, the comparative fit index (CFI), the Tucker–Lewis Index (TLI), the root mean square error of approximation (RMSEA) and the standardized root mean square residual (SRMR; Jackson, Gillaspy, & Pure-Stephenson, 2009). The $\chi^2/df$ was also considered. Factor loadings were assessed for statistical significance at the $p < .01$ level.

**Student Characteristics.** Each participating student provided information about gender, age, and grade level.

**Relational Aggression:** The five-item RA subscale from the Children’s Social Behavior Scale-Self Report (Crick & Grotpeter, 1995) was used. Each item (e.g., “Some kids tell lies about classmates so that the other kids won’t like the classmates anymore. How often do you do this?”) is scored on a 5-point scale ($1 = never$ to $5 = always$). This study relied on children’s reports of their own relational aggressive behavior, because children may have greater awareness of their own behavior than do parents, teachers, and peers (Lansford et al., 2012). In the same vein, by peer nomination assessment, it is more likely that apparently aggressive children may come easily to mind, whereas quieter and maybe more manipulative aggressors may not be rated as such (Archer & Coyne, 2005). Moreover, the present scale conceptualizes RA including behaviors that represent purposeful attempts to harm or threats to harm another’s peer relationships (Crick & Grotpeter, 1995). The scale has been used with American (Choi, Johnson, & Johnson, 2011; Terranova, Morris, & Boxer, 2008), Japanese (Kawabata, Crick, & Hamaguchi, 2010), and Greek samples (Kokkinos et al., 2014). Previous studies have shown adequate psychometric properties for this scale, namely, high internal reliability and validity (e.g., Kawabata et al., 2010; Kokkinos et al., 2014), whereas its unidimensionality with Greek samples has been established through CFA (Kokkinos et al., 2014). For the present data, CFA pointed out a single factor, $\chi^2 = 2.834$ ($df = 4, p = .000$), $\chi^2/df = .71$, and the indexes, CFI = 1.00, TLI = 1.03, SRMR = .025, RMSEA = .000. In addition, the scale’s Cronbach’s alpha coefficient was .70.

**Moral Disengagement.** The 14-item version of the MD scale specifically designed and validated for elementary school children (Caprara, Pastorelli, & Bandura, 1995) was used. According to Gini et al.’s (2014) meta-analytic review), this self-report scale has shown good reliability and is by far the most commonly used measure of MD across countries such as Italy and the Unites States, as only a few studies have used different assessment scales. The items describe an individual’s tendency to interpret harmful behavior as serving righteous purposes, to render activities benign by advantageous comparison, and to minimize the harmful effects of one’s inappropriate behavior (e.g., “To hit obnoxious classmates is just giving them ‘a lesson’”). Children rated the degree of their approval of MD on a 5-point scale ($1 = strongly disagree$ to $5 = strongly agree$). Although the items describe different mechanisms, previous studies using this scale with children (e.g., Caprara et al., 1995; Gini, 2006; Kokkinos et al., 2014) have demonstrated a unidimensional structure of the scale, which was replicated in this study using CFA, $\chi^2 = 93.600$, $df = 71$, $p = .000$, $\chi^2/df = 1.32$, CFI = .93, TLI = .91, SRMR = .065, RMSEA = .0.49. In this study, Cronbach’s alpha coefficient for the total scale was .79.

**Theory of Mind.** ToM was assessed by a ToM battery based on Wellman and Liu (2004) and Peterson, Wellman, and Slaughter’s (2012) scale tasks (i.e., first-order contents false-belief other task, hidden emotion, and sarcasm), along with four strange stories (White, Hill, Happé, & Frith, 2009) and three faux pas scenarios (Baron-Cohen, O’Riordan, Stone, Jones, & Plaisted, 1999). Both meta-analyses of Wellman, Cross, and Watson (2001) and Liu, Wellman, Tardif, and Sabbagh (2008) suggest that false-belief measures are reliable for the assessment of ToM skills. They have
been used across different cultures, such as European and non-Western, demonstrating adequate psychometric properties. In developing the procedural protocols prior to testing, back-translation from Greek to English was used to ensure that task narratives and questions were comparable to one another and had the same focus in both cultures (S. Mavropoulou, personal communication, May 18, 2013). According to Shahaeian, Peterson, Slaughter, and Wellman (2011), in their initial scale development, Wellman and Liu (2004) had carefully matched and counterbalanced all procedures, stimuli, and formats across tasks to reduce the likelihood that any differential executive, memory, linguistic demands, or cultural effects could plausibly account for observed differences in the relative difficulty of individual scale tasks.

**ToM Scale Tasks.** The three-item ToM scale includes a contents false-belief-other task, assessing children’s recognition that another person will predict the contents of a container on the basis of its outward appearance (Band Aids) rather than its actual contents (a toy pig); a hidden emotion task, assessing the protagonist’s inner psychological state and his deliberate intentions and active actions to deceive someone else (Peterson, Wellman, & Liu, 2005); and a sarcasm task, assessing the appreciation of a speaker’s communicative intent in nonliteral situations (Peterson et al., 2012). The last two tasks were modeled to guarantee suitable comparability in style and format to the other ToM scale tasks (see Peterson et al., 2012, for details). The first-order false-belief task was administered to ensure that participants had basic ToM skills before they attempted the more complex strange stories and faux pas tasks.

As in Peterson et al. (2005), correct responses to all control, as well as test questions, were required to pass any given task. The exception was the hidden emotion task, where Peterson et al.’s (2005) procedure was followed by using an additional explanation question and, consequently, using an alternative to the original scoring procedure (see Peterson et al., 2005, p. 517, for exact details).

Memory and reality control questions were used, and all control questions had to be passed in addition to the test questions for the child to be credited with passing the individual task. From the total of 140, 20 preadolescents (14.3%) failed to pass the false-belief task and therefore were excluded from the rest of the analyses. These results support Bosacki and Wilde Astington’s (1999) claim that in early adolescence, students’ success or failure in false-belief tasks could be explained by individual differences. Specifically, it is thought that the influences of factors, such as social competence, on mentalizing abilities become more prevalent in preadolescence than in late childhood (Calero, Salles, & Sigman, 2013). For each task that was passed, the child received 1 point, resulting in total possible scores ranging between 0 and 3. The gender of the story protagonists matched the child’s gender, and the order in which the tasks were presented was randomized and counterbalanced.

**Strange Stories.** Strange stories are short vignettes depicting social situations; each was followed by a single open-response question that required participants to explain a character’s behavior with reference to his or her mental states (see White et al., 2009, for the full text of each item and the coding scheme). Participants’ responses were coded using a 3-point scale, indicating full or partial or failed understanding (2/1/0). Specifically, a 3-point scale (0-2) assessed the degree to which the response referred to a mental state (2), to a fact (1), or to general, nonspecific information (0). The stories were selected on the basis of how demanding they were, based on Fletcher et al’s (1995) and White et al’s (2009) studies. These stories depicted a double bluff, a misunderstanding, a white lie, and, finally, a persuasion scenario. For the present study, the total score ranged from 0 to 8.

**Faux Pas Scenarios.** The faux pas task taps mental-state reasoning that is relevant to everyday peer interaction (Banerjee, Watling, & Caputi, 2011). The detection of a faux pas requires the appreciation that the knowledge of the speaker may be different from the listener’s and that the emotional impact of a statement may be different on the listener than the speaker (Baron-Cohen...
et al., 1999). Three scenarios were chosen from the battery of 10 faux pas stories (Baron-Cohen et al., 1999) based on cultural appropriateness. A similar selection of stories has been used in the study by Barlow, Qualter, and Stylianou (2010). After presenting each story, four questions (detection, identification, comprehension, and false belief) were asked to assess participants’ understanding. The coding system of Baron-Cohen et al. (1999) was applied. One point was assigned if all four questions were correctly answered. Failure of any of these four questions or even the detection question led to a zero score for that story. For the present study, the total score ranged from 0 to 3.

For all ToM tasks, inter-rater agreement was reached between the first author and a co-rater based on a sample of 20% of the answers to each question selected randomly for each task for strange stories and faux pas scenarios respectively (inter-rater correlation coefficients ranged between .79 and .81).

**Verbal Ability.** Because performance in the faux pas tasks has exhibited a positive correlation with verbal ability (Baron-Cohen et al., 1999), participants were assessed with the vocabulary subtest of the standardized Greek Wechsler Intelligence Scale for Children—Third Edition (Georgas, Paraskevopoulos, Bezevegis, & Giannitsas, 1997) to detect any severe delays in verbal ability. This subtest, consisting of 30 verbal items of increasing difficulty, is part of the verbal comprehension subscale and measures children’s knowledge of words and formation of concepts (Wechsler, 2003). It has shown excellent internal consistency, test–retest reliability, and inter-rater reliability, as well as construct validity by correlating significantly with other intelligence measures (Wechsler, 2003). An exclusion criterion of a score falling below 2 standard deviations from the class score was applied, ensuring that children who participated had at least a low average score of verbal ability (Wechsler, 2003). None of the participants was excluded from the analyses based on this criterion.

**Procedure**

Study participants were recruited by sending a letter to the school principals explaining the purpose and the procedure of the study, along with a consent form for the parents to be handed in by all pupils who did not present any reading difficulties according to the class teacher evaluations. Students were allowed 1 day to return the forms only if their parents did not wish them to participate in the project. The refusal rate was minimal (only 10 of a total of 150 parents contacted). The study was carried out in two sessions. During the first session, which lasted approximately 20 min, the paper-and-pencil questionnaires (i.e., MD and RA) were group administered to the participants, who also took a numbered card, which they were asked to keep until the second session for matching purposes. No names or other identifying information was requested.

The second session was conducted individually in a quiet room at school and lasted 15 to 20 min. Testing began with the administration of the 30-word vocabulary list and was followed by the ToM tasks. Each child was individually tested. All tasks were presented to each participant on a laptop screen using Windows Live Movie Maker software, both as written and spoken text to assist with comprehension. Each story was prerecorded by an adult female speaker and lasted 60 s. Participants were invited to follow the words while listening to the recording. Five hundred milliseconds after the spoken text had ended, a question about its content was presented on the screen and simultaneously spoken by the prerecorded voice. The experimenter recorded the child’s answer verbatim on paper to facilitate the scoring procedure. A similar procedure was applied by White et al. (2009). Tasks were presented with two different orders, which were randomly assigned across participants.
Table 1
Means, Standard Deviations, and Correlations among Relational Aggression, Moral Disengagement, Theory of Mind, and Verbal Ability

<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Relational aggression</td>
<td>1.44</td>
<td>.45</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Moral disengagement</td>
<td>2.00</td>
<td>.57</td>
<td>.31b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Total scale ToM</td>
<td>2.21</td>
<td>.65</td>
<td>.18a</td>
<td>.17</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Strange stories</td>
<td>5.35</td>
<td>1.7</td>
<td>.25b</td>
<td>.29b</td>
<td>.34b</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Faux pas</td>
<td>1.87</td>
<td>.99</td>
<td>.11</td>
<td>.26b</td>
<td>.19a</td>
<td>.24b</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Aggregate ToM score</td>
<td>9.11</td>
<td>2.70</td>
<td>.26b</td>
<td>.34b</td>
<td>.72b</td>
<td>.74b</td>
<td>.67b</td>
<td></td>
</tr>
<tr>
<td>7. WISC-III Vocabulary</td>
<td>30.67</td>
<td>5.91</td>
<td>.06</td>
<td>.30b</td>
<td>.32b</td>
<td>.32b</td>
<td>.30b</td>
<td>.44b</td>
</tr>
</tbody>
</table>

aCorrelation is significant at the .05 level (2-tailed).
bCorrelation is significant at the .01 level (2-tailed).

Overview of Statistical Analyses

Prior to the analyses, the relationships between conceptually overlapping measures were examined (i.e., ToM scale tasks, strange stories, and faux pas). Correlation matrices suggested that the three moderately correlated variables could be combined in an aggregate ToM score, which was created by averaging z scores of the three ToM measures. Bivariate correlations were then computed among all the variables used in the hypothesized mediation model to highlight significant relationships among the predictors, the outcome, and the mediators, which are prerequisite conditions for testing the mediation.

Moderated mediation analyses were performed on SPSS using the PROCESS macro (Hayes, 2013). The PROCESS moderated mediation analysis (Model 15) in this study included one independent variable (ToM or MD), one mediator (ToM or MD), one moderator (gender), and one dependent variable (RA). Verbal ability was treated as a covariate in both analyses. Bias-corrected bootstrapping procedures were used to test the moderated mediation models shown in Figures 1 and 2. The number of bootstrap samples for the bias-corrected bootstrap confidence intervals (CIs) was 5,000. Confidence intervals that do not contain zero indicate a significant indirect effect via the specific mediator. Data normality was checked using skewness, kurtosis values, and visual inspection of histograms and box plots.

RESULTS

Means, standard deviations, and Pearson bivariate correlations for all measures are shown in Table 1. Relational aggression was significantly positively correlated with MD. In addition, a significant negative correlation emerged between RA, total ToM scale score and strange stories, whereas faux pas was not correlated with RA.

Testing the Moderated Mediation Models

The models examined whether the indirect effect of MD or ToM on RA was mediated by ToM or MD and whether this mediation was moderated by gender.

With regard to Model 1, results showed a direct effect of ToM on RA for girls (B = -.09; 95% CIs [-.15, -.03]) but not for boys (B = .02; 95% CIs [-.05, .08]). Individual tests for mediation using bootstrap estimation of indirect effects with 5,000 replications confirmed the mediating role of MD in the association between ToM and RA only for boys (B = -.03, 95% CIs [-.08, -.008]), but not for girls.
Table 2

<table>
<thead>
<tr>
<th>Mediator variable model</th>
<th>B (SE)</th>
<th>T</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>MD</td>
<td>-.86 (.32)</td>
<td>-2.73</td>
<td>.01&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>WISC-III (covariate)</td>
<td>.13 (.03)</td>
<td>4.38</td>
<td>.00&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Outcome Variable: RA</th>
<th>B (SE)</th>
<th>T</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>ToM</td>
<td>.12 (.07)</td>
<td>1.77</td>
<td>.08</td>
</tr>
<tr>
<td>MD</td>
<td>.86 (.29)</td>
<td>3.00</td>
<td>.003&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Gender</td>
<td>.74 (.35)</td>
<td>2.10</td>
<td>.04&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>ToM × gender</td>
<td>-.11 (.04)</td>
<td>-2.52</td>
<td>.01&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>MD × gender</td>
<td>-.37 (.17)</td>
<td>-2.26</td>
<td>.03&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>WISC-III Vocabulary</td>
<td>.01 (.01)</td>
<td>1.41</td>
<td>.16</td>
</tr>
</tbody>
</table>

Note. MD = moral disengagement; ToM = theory of mind; RA = relational aggression; WISC-III = Wechsler Intelligence Scale for Children.

<sup>a</sup>.05.
<sup>b</sup>.01.

(B = -.01, 95% CIs [-.04, .004]). Together, these results support a partial mediation model, although the indirect effect of MD was marginal. Thus, regarding the indirect effects, ToM was posited to influence RA both directly (for girls) and indirectly through MD (for boys; Table 2).

In the case of Model 2, a direct effect of MD on RA for boys (B = .48, 95% CIs [.22, .75]), but not for girls (B = .11, 95% CIs [-.08, .30]) was in place. Individual tests for mediation using bootstrap estimation of indirect effects with 5,000 replications confirmed the mediating role of ToM in the association between MD and RA only for girls (B = .08, 95% CIs [.02, .21]), but not for boys (B = -.01, 95% CIs [-.11, .05]). Together, these results support a partial mediation model, although the indirect effect of ToM was marginal. Thus, regarding the indirect effects, MD was posited to influence RA both directly (for boys) and indirectly through ToM (for girls; Table 3).

**DISCUSSION**

Taking into account the recent advances highlighting the effects of social cognition on RA (e.g., Crain et al., 2005; Givens, 2010; Mathieson et al., 2011; Xie et al., 2002), the present study responded to a gap in the literature by identifying the mechanisms that underlie the interactions among two sociocognitive variables (ToM and MD) associated with RA using a moderated mediation analysis. As expected, both variables were significantly correlated with RA. Specifically, RA was negatively associated with ToM, contrary to H1. However, it is in agreement with previous findings with early adolescents indicating that aggressors show deficits in ToM skills compared with their peers and that they may be biased when they process social cues (Gomez-Garibello & Talwar, 2015; Shakoor et al., 2012). It seems therefore that students who engage in RA are more likely to show maladaptive sociocognitive functioning and use aggressive methods to solve conflicts (Caravita et al., 2010). Additionally, as it was expected (H2), MD was positively correlated with RA. This is in accordance with Bandura (2002), who argued that MD is linked with aggression proneness and
Table 3

**Moderated Mediation Analysis (Model 2)**

<table>
<thead>
<tr>
<th>Mediator variable model</th>
<th>B (SE)</th>
<th>T</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>ToM</td>
<td>-.07 (.03)</td>
<td>8.94</td>
<td>.00c</td>
</tr>
<tr>
<td>WISC-III (covariate)</td>
<td>-.02 (.01)</td>
<td>-1.95</td>
<td>.05a</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Outcome Variable: RA</th>
<th>B (SE)</th>
<th>T</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>MD</td>
<td>.86 (.29)</td>
<td>2.99</td>
<td>.003b</td>
</tr>
<tr>
<td>ToM</td>
<td>.12 (.07)</td>
<td>1.77</td>
<td>.08</td>
</tr>
<tr>
<td>Gender</td>
<td>.74 (.35)</td>
<td>2.10</td>
<td>.04a</td>
</tr>
<tr>
<td>MD × Gender</td>
<td>-.37 (.17)</td>
<td>-2.26</td>
<td>.03a</td>
</tr>
<tr>
<td>ToM × Gender</td>
<td>-.11 (.04)</td>
<td>-2.52</td>
<td>.01a</td>
</tr>
<tr>
<td>WISC-III Vocabulary</td>
<td>.01 (.01)</td>
<td>1.41</td>
<td>.16</td>
</tr>
</tbody>
</table>

Note. MD = moral disengagement; ToM = theory of mind; RA = relational aggression; WISC-III = Wechsler Intelligence Scale for Children.

a.05.  
b.01.  
c.00.

with similar research demonstrating positive associations between moral justification mechanisms and RA (Kokkinos et al., 2014; Murray-Close et al., 2006).

To gain more insight into the interactions between sociocognitive factors and RA, the present study tested two integrated models in which ToM and MD acted as mediators and gender as a moderator. In this line, the study yielded two key findings. First, MD partially mediated the relationship between ToM and RA. This implies that low levels of ToM indirectly, via the amplification of MD, have an effect on RA. Second, the mediating pathway of the association between ToM and RA through MD was moderated by gender, which also buffered the direct effect of ToM on RA. The present findings partially support H3 of the study and suggest that preadolescent boys with a deficient ToM may be more likely to morally disengage from their actions, which in turn results in RA. Possibly, this finding is a reflection of boys’ increased likelihood to engage in aggression when they possess a deficient ToM (Shakoor et al., 2012). The effect of ToM on MD is consistent with previous research showing that ToM judgments seem to affect moral cognition (Killen et al., 2011), as individuals’ beliefs about someone’s mind may precede the process by which they reach moral judgments (Knobe, 2005). Similarly, Baird and Astington (2004) found a negative correlation between false-belief tasks and children’s MD. The latter has also been found to bear a positive association with RA, especially among boys who are less likely than girls are to perceive RA as harmful for the victim (Murray-Close et al., 2006).

Relational aggression was also directly related to low levels of ToM only for girls, suggesting that girls with a deficient ToM were more likely to get involved in RA, a finding that was contrary to H5. The effect of poor ToM on RA only for girls is inconsistent with previous research suggesting that possessing good ToM skills may be linked to RA, especially among girls in middle childhood and early adolescence (Andreou, 2006; Espelage et al., 2004). However, this finding fits with evidence from research on school-aged children showing that deficits in ToM pose a greater risk for engagement in RA, as children tend to perceive RA as less harmful and cannot understand the
extent of damage it causes (Murray-Close et al., 2006). In the same line, Bosacki (2013) and Bosco, Gabbatore, and Tirassa (2014) found that preadolescent girls’ advanced ToM referred to more social and positive aspects of self and relationships, indicating that they are less likely to participate in aggressive behaviors.

As far as the second model is concerned, results showed that MD had a direct significant effect on RA but only for boys, suggesting that boys who use this strategy, which allows them to cognitively moralize their actions that would otherwise be considered against personal norms, may exhibit RA. Evidence from similar studies has been generally inconsistent regarding the role of gender in the link between high MD and RA (Gini et al., 2014; Kokkinos et al., 2014). Yet, the current finding is in agreement with those indicating that it is more possible an elevated use of MD to predict general aggression among boys (e.g., Gini et al., 2014; Paciello et al., 2008).

Interestingly, ToM was found to partially mediate the relationship between MD and RA, thus confirming the corresponding hypothesis (H4). In terms of gender, girls who used less MD possessed higher ToM skills, which resulted in less RA. This evidence is in accordance with other findings showing that girls are more likely than boys to consider RA as hurtful for the victims (Murray-Close et al., 2006) and as a result use less MD (Paciello et al., 2008). Thus, individuals may first judge a behavior as morally good or bad and then use this judgment to arrive at decisions about the mind (Knobe, 2005; Leslie et al., 2006). Consequently, the study supports that preadolescent girls’ low MD sets the basis for the activation of specific ToM skills that in turn lead to reduced involvement in RA (Shakoor et al., 2012).

**Practical Implications**

This study contributes to the existing literature by shedding further light on preadolescents’ participation in RA, as it is the first to use a moderated mediation analysis to investigate the possible mediators and moderators of RA among this age group and to evaluate the mechanisms underlying the development of these behaviors. The current findings offer a better understanding of the sociocognitive factors used by those who are at high risk for RA.

Results clearly indicated that relational aggressors appeared to be cognitively inferior (low ToM) and morally delayed, whereas cognitively skilled and emotionally sensitive preadolescents were less likely to be involved in RA. These findings have significant implications for developmental and school psychologists and point to a necessity to apply differential intervention strategies that should not solely rely on fostering strategic social competencies, but include programs of moral education (Gasser & Keller, 2009; Nucci, 2001). Indeed, there is a lot of evidence supporting the use of social–cognitive reframing interventions with relationally aggressive youth (Fraser et al., 2005; Leff et al., 2010; Voulgaridou & Kokkinos, 2015). However, the suggestion that aggressors should be trained to improve their ToM skills should be treated with caution, as perpetrators form a notoriously heterogeneous group, and it is possible that no “blanket” technique will be completely effective (Sutton et al., 1999). According to Nucci and Narvaez (2008), educational interventions aimed at fostering both moral resilience and social competences in children have been shown to be more effective.

Regarding relational aggressors’ low ToM, Caravita et al. (2010) proposed that interventions should include actions such as interpreting ambiguous social cues as less hostile or choosing nonaggressive strategies to solve conflict situations. Moreover, intervening programs that include students’ training on cognitive skills regarding the understanding of others’ emotions, perceptions, and beliefs are key factors in the weakening of aggressive behavior (Kerig, 2007; Kokkinos & Kipritsi, 2012). Employing training strategies for children with poor ToM in their schooling years may help improve their social interactions. For instance, evidence supports that discussing scenarios
of false-belief and mental states improves children’s understanding and explaining of other’s behavior (Shakoor et al., 2012).

Given the findings of the current study, future educational interventions that promote the development of children’s social competence may want to incorporate and/or more strongly emphasize efforts to facilitate moral emotions of relational aggressors (Malti, Gasser, & Gutzwiller-Helfenfinger, 2010). In this regard, educational interventions against RA would be benefited by emphasizing “moral training” to target immoral conceptions. Specifically, strategies could focus on rendering relational aggressors capable of understanding the harm their aggressive behavior causes to the victims. Furthermore, interventions should emphasize that participation in RA is morally reprehensible and that judgments of relationally aggressive behavior should include a concern for fairness and the welfare of others (Murray-Close et al., 2006). Parents and teachers could be employed to guide children’s immoral thoughts in moral engagement and to assist them with recognizing and taking responsibility for their actions (Gini et al., 2011).

The findings also indicated that the intervention programs should be differentiated according to gender. Boys’ higher levels of MD and girls’ poor ToM are associated with greater participation in RA. As boys believe that RA is relatively acceptable and are more likely to morally disengage from these acts, they may benefit from an increased emphasis on interpersonal relationships and caring behavior skills training (Murray-Close et al., 2006). In line with this reasoning, Bosacki (2013) proposes interventions that enhance socioemotional and cognitive development in early adolescent girls aimed at increasing their ToM and consequently engage in prosocial behaviors.

Limitations and Future Research

This study has several limitations that should be taken into consideration when interpreting the results. First, its cross-sectional design limits its ability to confirm a causal relationship between the variables under investigation. Thus, the results should be interpreted carefully, whereas future longitudinal designs would be able to clarify the prospective relationships among the variables. Second, the present research used self-report measures of RA and MD. Even though the scales have been validated, they may underestimate or overestimate the actual degree of these behaviors reported by the participants. Hence, future studies would provide more accurate information regarding the complex mechanisms underlying the relationship between sociocognitive factors and RA by using standardized interviews or observational methods. Third, the sample included only preadolescents. Therefore, the proposed mediational pathway could be examined in other age groups, such as children and adolescents. A final concern is that no information about victims was obtained.

Overall, the study extends previous literature by placing emphasis on the simultaneous effects of sociocognitive factors influencing RA moderated by gender. Results indicated that both the direct and indirect effects of sociocognitive factors were found to be a function of gender. Specifically, MD mediated the relationship between ToM and RA only for boys. Also, morally disengaged justifications seemed to contribute to deficit ToM skills, which in turn led to RA only for girls. Additionally, MD had a direct effect on RA for boys, whereas ToM was found to directly affect RA only for girls. However, more research is needed to further clarify and elaborate on the present findings to offer a more comprehensive picture of the models in this study by including contextual variables.

REFERENCES


Psychology in the Schools DOI: 10.1002/pits


Psychology in the Schools DOI: 10.1002/pits


Psychology in the Schools DOI: 10.1002/pits