Intrapersonal Risk Factors of Aggressive Behavior in Childhood: A Longitudinal Perspective

by

Fabian Kirsch

Doctoral Dissertation

submitted for the Degree of

Doctor of Philosophy

to the

Faculty of Human Sciences

University of Potsdam

Potsdam, June 2017

Erstbetreuerin: Prof. Dr. Barbara Krahé
Zweitbetreuer: Dr. Robert Busching

Erstgutachterin: PD Dr. Daniela Niesta-Kayser
Zweitgutachterin: Prof. Dr. Nina Heinrichs

Datum der Verteidigung: 8. November 2017

Published online at the
Institutional Repository of the University of Potsdam:
URN urn:nbn:de:kobv:517-opus4-407369
http://nbn-resolving.de/urn:nbn:de:kobv:517-opus4-407369

Danksagung

Während der Arbeit an der vorliegenden Dissertation wurde ich von zahlreichen Personen methodisch, emotional und motivational unterstützt. Ich möchte mich daher an dieser Stelle ganz herzlich bei ihnen bedanken:

- Prof. Dr. Barbara Krahé für die Möglichkeit unter ihrer Erstbetreuung meine Promotion schreiben zu dürfen, für die Zeit, die sie sich stets für mich genommen hat, und für ihre wissenschaftliche und methodische Unterstützung
- Robert für die Zweitbetreuung meiner Arbeit und für seine Impulse, mich methodisch weiter zu entwickeln
- Daniela für ihr stets offenes Ohr für meine Bedenken und Probleme, und für ihre bedingungslose Bereitschaft, mich zu unterstützen und zu fördern
- Greta für ihre fachliche, emotionale und motivationale Unterstützung in allen Lebenslagen und für zahlreiche tolle Mittagessen, Zugfahrten und Telefonate
- Helena für die stets angenehme und erfolgreiche Zusammenarbeit an insgesamt fünf gemeinsamen Papern
- das gesamte ehemalige und noch bestehende PIER-Team; insbesondere meine ausnahmslos tollen Mitdoktorand/innen, die wir in den letzten Jahren gemeinsam viele Umbrüche und Veränderungen innerhalb des Projekts erlebt und gemeistert haben: Ania, Anna, Finja, Greta, Johannes, Nele, Sebastian, Susanne und Svenja
- meine bisher nicht genannten ehemaligen und aktuellen Kolleg/innen der Abteilung Sozialpsychologie für die angenehme Arbeitsatmosphäre: Ariane, Brigitta, Evi, Isabell, Jana, Janis, Jeanette, Johannes, Juliette, Lisa, Nicole, Paulina, Robin, Ronja und Sophie
- Lara für ihre tatkräftige Unterstützung bei der Erhebung der Attraktivitätsdaten
- Melanie für das gründliche Korrekturlesen dieser Dissertation
- meinem gesamten Freundeskreis für die nötige Ablenkung bei diversen Spieleabenden
- meinen Eltern für ihre moralische Unterstützung
- meinem Freund Gregor für seine unerschöpfliche Geduld mit all meinen Hochs und Tiefs im Laufe der Promotion

The present doctoral dissertation is based on following empirical papers:

(1) Kirsch, F., Busching, R., Rohlf, H., & Krahé, B. (2017). Using behavioral observation for the longitudinal study of anger regulation in middle childhood. *Applied Developmental Science*. Advance online publication. doi:10.1080/10888691.2017.1325325

(2) Holl, A. K., Kirsch, F., Rohlf, H., Krahé, B., & Elsner, B. (2018). Longitudinal reciprocity between theory of mind and aggression in middle childhood. *International Journal of Behavioral Development*, 42, 257-266. doi:10.1177/0165025417727875

(3) Kirsch, F., Busching, R., Rohlf, H., & Krahé, B. (under review). Physical attractiveness, peer problems and aggressive behavior in childhood: A three-year longitudinal study.

Table of Contents

Abstract

Background: Aggressive behavior at an early age is linked to a broad range of psychosocial problems in later life. That is why risk factors of the occurrence and the development of aggression have been examined for a long time in psychological science. The present doctoral dissertation aims to expand this research by investigating risk factors in three intrapersonal domains using the prominent social-information processing approach by Crick and Dodge (1994) as a framework model. Anger regulation was examined as an affective, theory of mind as a cognitive, and physical attractiveness as an appearance-related developmental factor of aggression in middle childhood. An additional goal of this work was to develop and validate a behavioral observation assessment of anger regulation as past research lacked in ecologically valid measures of anger regulation that are applicable for longitudinal studies.

Methods: Three empirical studies address the aforementioned intrapersonal risk factors. In each study, data from the PIER-project were used, a three-wave-longitudinal study covering three years with a total sample size of 1,657 children in the age between 6 and 11 years (at the first measurement point). The central constructs were assessed via teacher-reports (aggression), behavioral observation (anger regulation), computer tests (theory of mind), and independent ratings (physical attractiveness). The predictive value of each proposed risk factor for the development of aggressive behavior was examined via structural equation modeling.

Results and Conclusion: The newly developed behavioral observation measure was found to be a reliable and valid tool to assess anger regulation in middle childhood, but limited in capturing a full range of relevant regulation strategies. That might be the reason, why maladaptive anger regulation was not found to function as a risk factor of subsequent aggressive behavior. However, children's deficits in theory of mind and a low level in physical attractiveness significantly predicted later aggression. Problematic peer relationships were identified as underlying the link between low attractiveness and aggression. Thus, fostering children's skills in theory of mind and their ability to call existing beliefs about the nature of more versus less attractive individuals into question may be important starting points for the prevention of aggressive behavior in middle childhood.

List of Tables

Paper 1

Paper 2

Paper 3

List of Figures

Paper 1

Paper 2

Paper 3

Appendix

1

Introduction

Aggression is one of the most worrisome behaviors a human being can show. On a daily basis, we are confronted with the experience of aggressive acts or stories and reports about them; for example, in the news, social media, movies, or TV series. Severe acts of aggression include terrorism, rape, torture, and war. However, "milder" acts of aggression may be likewise disturbing for the receiver of that behavior, be it getting physically hurt or being the victim of rumors or social exclusion. Based on this reasoning, it becomes very clear that psychological science – social and developmental psychology in particular – has investigated aggressive behavior for a long time. In this context, the prediction of aggressive behavior is one of the most relevant aspects to find approaches that reduce the occurrence of aggressive acts on an individual or group level. Since childhood is a formative time of an individual's life, it seems considerably important to examine factors that contribute to the development of aggression already from an early age.

The present doctoral dissertation aims to identify risk factors of aggressive behavior in middle childhood. The term *middle childhood* usually refers to primary school age and, thus, includes the age from 6 to 12 years (Collins, 1984). In this work, the focus is on *intra*personal risk factors of aggressive behavior, that is, factors that are inherently attached to the individual. Three domains of intrapersonal risk factors are considered, namely children's affective, cognitive, and appearance-related characteristics. Specifically, children's impaired ability to adaptively regulate the emotion of anger is examined as an affective risk factor. Furthermore, a child's deficits in attributing mental states to others (i.e., theory of mind) are examined as a cognitive risk factor. Finally, a child's lack of physical attractiveness is examined as a distal appearance-related risk factor that may influence the development of aggression through social consequences, such as more versus less problematic peer relationships.

The doctoral dissertation starts with the theoretical background in *Chapter 2*. First, this chapter provides an overview of the development and prediction of aggressive behavior in children. Second, the three aforementioned risk factors that are in the focus of the present work are introduced. In this context, the possible links to the development of aggression are elaborated in terms of a superordinate framework model.

Based on the theoretical background, *Chapter 3* presents the specific aims of the doctoral dissertation and the corresponding empirical papers. *Chapter 4* introduces the general research concept that is shared by all empirical papers. Here, the total sample and procedures of data collection are summarized, followed by an overview of the central instruments and statistical strategies that were applied to test the hypotheses.

Chapter 5 deals with the empirical papers. Paper 1 presents a validation study of a behavioral measure of anger regulation, which includes the association to aggressive behavior. Paper 2 documents reciprocal effects between theory of mind and physical as well as relational aggression in a longitudinal study. Paper 3 demonstrates prospective effects of physical attractiveness on subsequent peer problems and aggression. The doctoral dissertation concludes with a general discussion of the results, general strengths and limitations of the empirical studies, and implications that can be derived from the findings (*Chapter 6*).

2

Theoretical Background

2.1 Aggressive Behavior

It depends on a variety of aspects to which behavior we refer to as aggressive, including culture, context, or personal point of view. However, to investigate the development of aggressive behavior scientifically it is important to describe this construct as precisely and unambiguously as possible. In social psychology there is a consensus to define aggressive behavior as "any form of behavior directed toward the goal of harming or injuring another living being who is motivated to avoid such treatment" (Baron & Richardson, 1994, p. 7).

The intention to harm, independent of the actual occurrence of harm, is central to this definition. Thus, harm that results from coincidence or carelessness is not referred to as aggression. Furthermore, the definition excludes any harmful acts which happen with the recipient's consent, such as sadomasochism or assisted suicide. On the other hand, this definition includes not only behavior that is intended to result in physical pain, but also more diverse and covert forms of aggression, as described in the following.

2.1.1 Physical versus Relational Aggression

Depending on the point of view, aggressive behavior can be classified in different subtypes (for a review, see Krahé, 2013). For example, aggression can be distinguished based on its function (reactive vs. proactive), visibility (covert vs. overt), or severity (transient vs. lasting effects). Central to this work is the differentiation in forms of occurrences, in particular the distinction between physically versus relationally aggressive behavior.

Physical aggression refers to behavior with the intention to induce physical harm to another individual, for example by hitting, punching, or kicking. Physical aggression is the first form of aggression that can be observed in human beings. Already toddlers show precursors of physical aggression, for example by biting or pulling one's hair. Many studies have shown that – from early childhood on – boys are more often involved in acts of physical aggression than girls (for a meta-analytic review, see Archer, 2004). This sex difference is usually explained by social learning: Since physically aggressive acts are viewed as masculine, boys, but not girls, may learn that using physical aggression can be appropriate in terms of a masculine sex role (e.g., Archer, 2004).

Relational aggression, in contrast to physical aggression, refers to behavior with the intention to damage social relationships of another individual, for example by spreading rumors, exclusion from group activities, or ignoring. There are several other terms describing this form

of aggression, such as *social aggression* or *indirect aggression*. Even though the definitions may differ slightly, all terms essentially refer to relational aggression (Archer & Coyne, 2005). Thus, for the sake of consistency, the term relational aggression is used throughout this work. Regarding sex differences, studies revealed mixed evidence. Some studies reported that girls show more relational aggression than boys do (e.g., Crick, 1997). Other studies, however, found no consistent difference between girls and boys (e.g., Lansford et al., 2012). A meta-analysis by Card, Stucky, Sawalani, and Little (2008) concluded that girls seem to exhibit more relational aggression than boys do, but that this sex difference is of trivial effect size. Differences in friendship structures may explain the potential distinction in their behavior: Compared to boys, girls are more interested and deeper involved in close, intimate peer relationships, which may fuel relationally aggressive behavior (Murray-Close, Ostrov, & Crick, 2007).

2.1.2 Development of Aggressive Behavior

To a specific extent, aggression is supposed to follow a normative developmental course. In the first two years after birth, physical forms of aggressive behavior increase, for example in the form of biting and pulling hair (Alink et al., 2006; Nærde, Ogden, Janson, & Zachrisson, 2014). However, from the fourth year of life, physical aggression decreases in its rate of occurrence (Alink et al., 2006). In contrast, due to enhanced linguistic competence, relational aggression emerges. For example, a child may threaten a friend to exclude him or her from a birthday party, if he or she is not willing to obey (Bonica, Arnold, Fisher, Zeljo, & Yershova, 2003). During school age, when peer relationships gain in importance (e.g., Lam, McHale, & Crouter, 2014), physical aggression is still declining (e.g., NICHD Early Child Care Research Network, 2004), whereas relational aggression tends to increase, especially in girls (Murray-Close et al., 2007). In addition, at this age, more sophisticated and complex forms of relational aggression develop, such as ignoring a peer by pretending not to see him or her (for a review, see Crick et al., 1999).

In the course of a lifetime, aggressive behavior seems to be considerably stable. Olweus (1979) reported a stability coefficient of aggression in males that is comparable to the stability of intelligence. In another study, Huesmann, Dubow, and Boxer (2009) examined the stability of aggression over a period of 40 years, starting at the age of 8 years. In total, they found an overall stability of moderate size, higher stabilities for males than for females, and higher stabilities in adulthood than in childhood or adolescence. Furthermore, Moffitt (1993, 2007) distinguished two prototypical patterns of aggression development: On the one hand, there is

antisocial behavior that persists over the lifetime, and is associated with unfavorable developmental and environmental conditions, such as frequent family conflicts. On the other hand, there is antisocial behavior of which the occurrence is limited to adolescence and which is assumed to be a reaction to the experience of discrepancy between biological and social age (i.e., feeling like an adult, but not being treated as such; *maturity gap*). Individuals that belong to the persistent aggressive group are at a higher risk for various maladjustments in later life, including depression, inferior educational attainment, low self-reported health, and an increased probability to get arrested (Huesmann et al., 2009). In line with this research, a meta-analysis by Card et al. (2008) revealed that the involvement in aggressive behavior is associated with peer problems and delinquent behavior. These results stress the relevance of psychological research to predict and prevent the development of non-normative levels of aggression.

To date, there are numerous studies examining potential risk factors for the development of aggressive behavior. For example, aggression can be predicted by problems in peer relationships (for a meta-analytic review, see Reijntjes et al., 2011), by the experience of maltreatment (e.g., Manly, Kim, Rogosch, & Cicchetti, 2001), and by the consumption of violent media content (for a review, see Krahé, 2016). In her review, Reebye (2005) concluded that there is no single pathway that can determine the development of aggressive behavior. Instead, there are several pathways that may function as the origin of an individual's aggressive behavior, including family dynamics, exposure to aggression, and intrapersonal factors.

2.1.3 Social Information Processing Model

Several biological and psychological theories aim to explain why individuals engage in aggressive behavior (for a review, see Krahé, 2013). A prominent theoretical approach is the *Social Information Processing* (SIP) Model by Crick and Dodge (1994), which lays the foundation for the present doctoral dissertation. According to the SIP Model, the emergence of social behavior in an isolated social situation, including aggression in childhood, is based on the processing of social information. The authors propose six steps that connect the initial occurrence of a situation with the final behavior (see Figure 1). Step 1 describes the perception and encoding of social cues, such as facial expressions, gestures, or verbal utterances. In step 2, these social cues are interpreted. Here, inferences about the sender's intentions are made and attributional processes take place. Next, in step 3, the receiver clarifies his or her goals. For example, he or she might be interested in maintaining a good relationship with the sender. In step 4, potential response options are contemplated that are beneficial (or not) to the previously

identified goals. Then, in step 5, based on an evaluation of the options, a decision for one response option is made. Finally, step 6 describes the performance of the chosen behavior. This behavior, in turn, is a social cue that triggers the SIP steps in the interaction partner. In this sense, SIP is constantly taking place in a social situation.

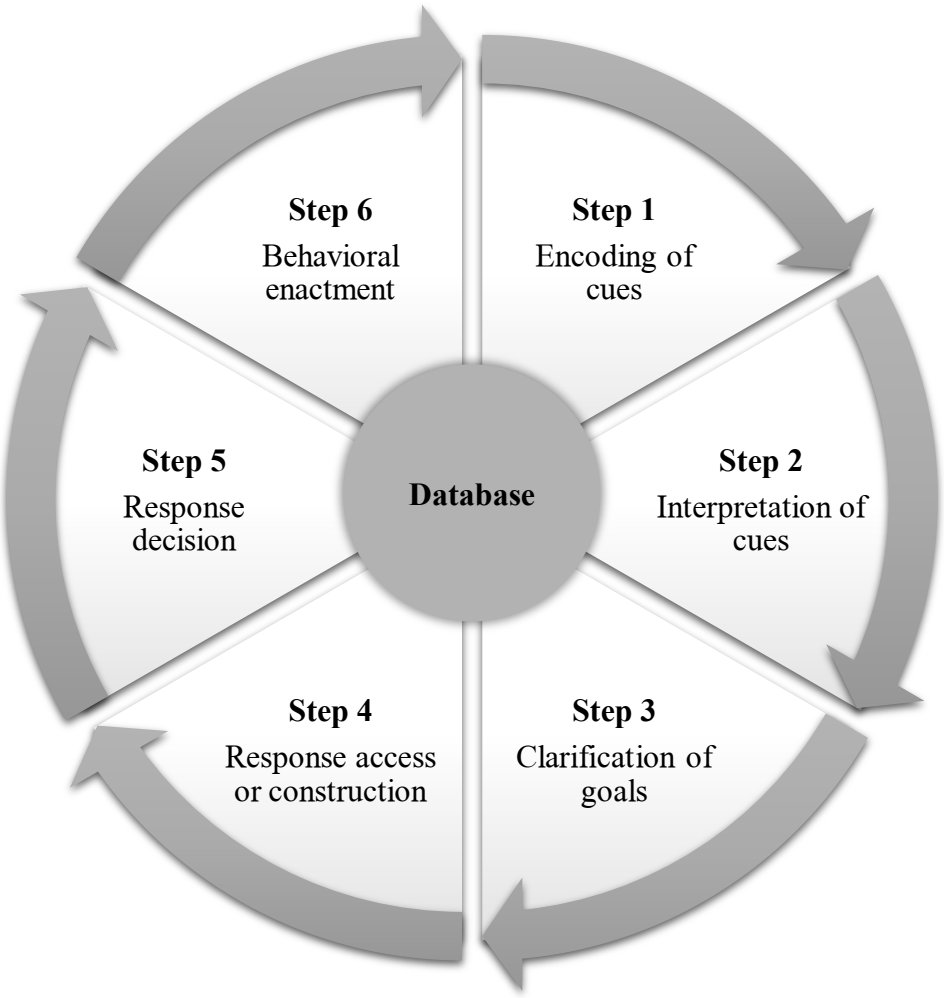

Figure 1. Simplified presentation of the Social Information Processing Model by Crick and Dodge (1994)

The following example of a social interaction between child A and child B may function as an illustration of the aforementioned steps: Imagine child B that is laughing after child A has stumbled. As a first step, A has to notice B's behavior; that is, perceiving B's laughter (step 1: encoding of cues). Next, A is interpreting B's behavior and may assume an underlying hostile intent (step 2: interpretation of cues). As a goal for this specific social situation, A may aim to demonstrate to B that he or she is no coward (step 3: goal setting). To achieve this goal, A collects several response options, for example offending B or laughing with B (step 4: response

access or construction). These optional responses are evaluated, and A may decide to laugh with B rather than to offend him or her (step 5: response decision). Finally, child A shows the chosen behavior, that is, A is laughing with B (step 6: behavioral enactment). This behavior in turn triggers the proposed SIP steps in child B. Crick and Dodge (1994) noted that, even the postulated steps follow a logical sequence, they can occur simultaneously and can feed back to previous steps (e.g., interpretation of cues may influence the encoding of further incoming cues within the same social situation).

A central aspect of the SIP model is the *database* (see Figure 1). It refers to the individual configuration of a person's memory of past social events. Therefore, the database includes social scripts and social knowledge. In all steps of the SIP model, the database is supposed to play an influential role. For example, during encoding (step 1), the database can selectively guide the attention to specific social cues. The interpretation of the perceived cues (step 2) can be influenced by social knowledge (e.g., child A is used to experience humiliation by his or her peers). In step 3, the clarification of goals can be influenced by social scripts (e.g., child A assumes that demonstrating his or her own strength is a normative goal). Next, memories of similar social situations can be contemplated to collect possible response options (step 4), and to decide on one response (step 5; e.g., child A remembers that laughing about one's own misfortunes usually feels less embarrassing). Thus, the database is important for virtually all SIP steps and, in consequence, for the occurrence and development of aggressive behavior.

As the database contains memories of past social events, social experiences can influence the database. That is, a child may experience less favorable peer interactions, for example, caused by his or her physically unpleasing appearance. These social experiences are stored in the database, which in turn influence subsequent social interactions. In an analogous manner, the involvement in aggressive behavior can cause unfavorable social experiences that are stored in the database. As a result, reciprocal relations between aggression and SIP can be expected. In one study, Lansford, Malone, Dodge, Pettit, and Bates (2010) demonstrated evidence for this assumption. They longitudinally studied children from kindergarten to third grade, covering four points of measurement. The authors found a reciprocal relation between aggression and SIP problems mediated by social preference over time. On the one hand, the children that were characterized by more SIP problems experienced less social preference, which in turn led to more aggressive behavior. On the other hand, the children that were aggressive experienced less social preference, which in turn led to more SIP problems.

In past research, the SIP model was often successfully used to explain and predict aggressive behavior. For example, studies showed that aggressive in comparison to nonaggressive children tend to attribute more hostile intentions to their peers in ambiguous social situations (step 2; for a meta-analytic review, see de Castro, Veerman, Koops, Bosch, & Monshouwer, 2002). In addition, they generate more aggressive responses and evaluate them less negatively (step 4 and 5; de Castro, Merk, Koops, Veerman, & Bosch, 2005).

Since SIP incorporates cognitive processes (e.g., attributional processes, decision making), individual differences in cognitive abilities can have an impact on the development of aggression; in particular, differences in social cognition (e.g., theory of mind). In addition, as emotions are known to influence cognitive processes, affective states can shape the SIP and the resulting social behavior. Thus, individual differences in emotional experience and regulation can also have an impact on the development of aggression (e.g., the regulation of anger). Finally, as described above, the database is supposed to influence SIP and inherently differs between individuals. Individual differences that may predict the probability for particular social experiences (that are stored in the database) can have an impact on the development of aggression as well (e.g., physical attractiveness).

2.1.4 Summary: Aggressive Behavior

Aggressive behavior describes behavior with the intention to harm another individual (Baron & Richardson, 1994), be it in a physical or a relational manner. Although some aggressive behavior is observable in virtually all children, research also suggests that persistent engagement in aggressive behavior is associated with a broad range of negative outcomes, including peer problems and internalizing problems (e.g., Huesmann et al., 2009). Therefore, it seems of considerable importance to identify factors that contribute to the development of aggression. Psychological research already examined a variety of potential risk factors, and a collection of theoretical explanations exist. A prominent approach is the SIP model by Crick and Dodge (1994) that is often used to explain the occurrence and development of aggression in childhood. Accordingly, aggressive behavior in a specific social situation is viewed as a result of a sequence of steps in which social information is processed. Several factors may influence this processing and, consequently, the probability of aggression; for example, by guiding the interpretation of an ambiguous situation, or by recollecting past similar events from the database. In the following chapters, the SIP model functions as a framework model for the prediction of aggressive behavior in childhood. Both cognitive, affective, and appearance-

related individual differences can be integrated in this model as they can influence specific SIP steps and the database, respectively. The specific associations between the potential intrapersonal risk factors, the SIP model, and aggression are discussed in the corresponding subchapters.

2.2 Emotion and Anger Regulation

Emotions are immanent in human life and have diverse functions for dealing with everyday requirements. For instance, emotions are important for decision making (Bechara, 2004), they serve as a source of motivation (Savolainen, 2014), and they can influence impression formation (Forgas, 2011). Despite their benefits, it is important to restrict the extent emotions are openly expressed and influence our behavior in modern society ("politics of emotions"; Shields, 2005). That is, which form and what intensity of emotion is socially appropriate depends on the given context. In this sense, emotions are required to be regulated in everyday life.

Emotion regulation "refers to the processes by which individuals influence which emotions they have, when they have them, and how they experience and express these emotions" (Gross, 1998, p. 275). Accordingly, emotion regulation influences the form, the occurrence, the intensity, and the expression of emotions. For example, imagine an 8-year-old boy who is used to be the best in class, but now he receives a bad grade in a math test. How will he feel? How intense will he experience his emotion? And to what extent is he going to express his emotions openly in front of class? All these questions refer to emotion regulation processes. In the last few decades, research on emotion regulation has gained a lot of attention (Adrian, Zeman, & Veits, 2011).

Among all emotional states, anger is considered to be one of the most difficult regulatable emotions, especially in childhood (e.g., Waters & Thompson, 2014). Anger is a negatively valenced emotion, which emerges when the individual's goals are blocked (for a review, see Potegal, Stemmler, & Spielberger, 2010). In terms of intensity, it can range from mild annoyance to strong rage. Early research conceptualized anger as a basic emotion, an emotion that is expressed and understood in a similar way across cultures (Ekman & Friesen, 1971). Anger activates the approach motivation system; that is, the individual tends to approach the anger-eliciting stimulus rather than to avoid this stimulus. Therefore, on a physiological level, anger is linked to increased blood pressure, arousal, and alertness. On a behavioral level,

anger is closely linked to aggression. Especially in childhood, anger is experienced as a common and intense emotion (von Salisch, 2000).

The term anger regulation is basically a specific differentiation of the superordinate term emotion regulation. In line with the definition by Gross (1998), anger regulation describes processes that influence the occurrence, intensity and expression of anger. For example, if the bad grade in the math test triggers anger in the 8-year-old boy, processes will be activated that influence whether the boy experiences anger immediately (occurrence), the degree of his anger (intensity), and to what extent he expresses his anger openly (expression).

Anger regulation in particular and emotion regulation in general can vary in terms of efficiency. Basically, regulatory efforts can be described as adaptive versus maladaptive (Gross & Thompson, 2007). Adaptive regulation is considered to be effective. That is, for example, negative emotions, such as anger, are efficiently decreased through regulatory processes. In contrast, maladaptive regulation is considered to be ineffective. That is, for example, regulatory attempts lead to a stabilization or increase rather than a decrease of anger. Cole and Hall (2008) further described maladaptive emotion regulation as the use of strategies that are ineffective in general or in the given context (e.g., trying to change an uncontrollable situation), that are socially inappropriate (e.g., laughing at a funeral), or that inhibit functional development (e.g., avoiding normative, but anxiety inducing situations).

In the literature, many regulation strategies are listed, on both a cognitive and a behavioral level. The classification of adaptive versus maladaptive strategies depends on the specific situation in which they are used (Gross & Thompson, 2007). However, in their validated questionnaire for assessing emotion regulation, Grob and Smolenski (2005) list strategies that are considered to be generally adaptive and maladaptive for several emotions, including anger. Adaptive strategies include problem-oriented behavior, distraction, and cognitive reappraisal. Maladaptive strategies include resignation, perseveration (cognitive focus on frustrating stimuli), and devaluation of the self. For instance, if the aforementioned boy ruminates over his bad math grade or instantly attributes the failure to his lacking math skills, his experienced anger will probably stabilize or even increase (maladaptive anger regulation). However, if he starts thinking about his past successes or attributes the bad grade to the high test difficulty, he will probably become less angry (adaptive anger regulation).

2.2.1 Development of Emotion Regulation

After birth, infants strongly depend on their parents when it comes to the regulation of their emotional experiences because they possess no or only few regulative skills (for a review, see Brenner & Salovey, 1997; Dimitrova & Lüdmann, 2014). For instance, parents usually calm down their crying or fearful infants by singing a song or lulling them to sleep. One of the first regulation strategies that emerge in infants is distraction, for example by looking away from the distressing stimuli (Crockenberg & Leerkes, 2004; Ekas, Lickenbrock, & Braungart-Rieker, 2013). With increasing linguistic and motoric skills, the possibilities for emotion regulation also increase. For example, children in preschool age are able to verbally share and discuss their discomfort in a given situation (Kopp, 1992). With school entrance, internal emotion regulation strategies increase. For instance, children start to use more cognitive strategies, such as cognitive distraction (Altshuler & Ruble, 1989). While growing up, a repertoire of various behavioral and cognitive regulation strategies develops (Brenner & Salovey, 1997). Thus, emotion regulation is primarily a function of age, and this is not only limited to infancy and childhood, but takes throughout the life span (e.g., Gross et al., 1997; Phillips, Henry, Hosie, & Milne, 2006). Basically, emotion regulation becomes more adaptive with increasing age.

Besides age there are other predictors of emotion regulation. For example, Otterpohl, Imort, Lohaus, and Heinrichs (2012) found that dysfunctional parental practices (i.e., punitive reactions in response to children's anger) are associated with maladaptive anger regulation in 3- to 11-year-old children. Moreover, which specific regulation strategies are used in an emotion-eliciting situation depends on what is thought to be socially accepted (Raval, Martini, & Raval, 2007). In addition, the development of emotion regulation can be shaped by individual differences, such as the children's temperament (Tan, Armstrong, & Cole, 2013). Although emotion regulation is developing with age and can be influenced by several factors, it is conceptualized as an individual difference variable (Cole, Michel, & Teti, 1994). That is, the adaptivity of emotion regulation is assumed to be relatively stable over time.

2.2.2 Measuring Anger Regulation

To date, assessing emotion regulation in general and anger regulation in particular is a challenging task (Adrian et al., 2011). One of the most widely used measures are children's self-reports about the use of their anger regulation strategies (Aldao, Nolen-Hoeksema, & Schweizer, 2010). Self-reports clearly have advantages. For example, they are easy to

administer and usually require only a short amount of time. However, in her review, Underwood (1997) highlighted the use of self-reports in the context of children's emotion regulation as problematic in several perspectives. One of the most pressing problems is that children's self-reports about their use of emotion regulation strategies can remarkably differ from their use in actual emotion-eliciting situations. For example, in a study by Parker et al. (2001), 6- to 11-year-old children mentioned more strategies to regulate their anger in hypothetical situations than they generated in actual anger-eliciting situations. Underwood (1997) argued that such a bias in self-reports can occur because children are not aware of their actual emotions. In addition, the impact of emotions on the use of regulation strategies may be underestimated. Emotional states are assumed to influence information processing (Lemerise & Arsenio, 2000), which in turn can influence regulatory behavior. In line with this reasoning, self-reports may be a useful tool to measure children's theoretical knowledge of emotion regulation, but it is unlikely that these reports directly correspond to children's actual behavior (Underwood, 1997).

To address this methodological problem, some researchers developed structured observational measures of emotion regulation in general and anger regulation in particular (e.g., Helmsen & Petermann, 2010; Rohlf & Krahé, 2015). Usually, they put the children into an actual emotion-eliciting situation and record their use of emotion regulation strategies. As emotion regulation is examined in a real-life context, this approach is assumed to reflect habitual behavior in natural situations (Cummings, Davies, & Campbell, 2000), even though the observation is limited to a single, short-term situation and to mere observable behavior (i.e., internal regulation cannot be assessed).

Central to the present doctoral dissertation is the behavioral observation measure developed by Rohlf and Krahé (2015). The authors created an anger-eliciting situation for 6- to 10-year-old children and observed which regulation strategies the children use to deal with their anger. The children were given ten wooden blocks and a photo that demonstrated how exactly the blocks should be stacked up to form a tower. The children were given roughly three minutes (2:40 min) to complete this task. However, unknown to the children, two of the wooden blocks were rounded on one side, which made the task extremely challenging and virtually impossible to complete in the given time. Rohlf and Krahé (2015) further developed a coding system to categorize children's use of anger regulation strategies. Specifically, they recorded the use of five strategies that can be classified as adaptive or maladaptive. The maladaptive strategies included the visual and verbal focus on the frustrating stimuli, resignation, and venting the anger. Accordingly, solution-oriented behavior was recorded as an adaptive regulation strategy.

The authors found this assessment to be a reliable and valid measure of anger regulation by correlating it with the conceptually similar construct of anger reactivity (construct validity) and with aggressive behavior and peer problems (criterion validity). In a follow-up study, this assessment was found to have predictive validity (Kirsch, Rohlf, & Krahé, 2015).

2.2.3 Anger Regulation and Aggressive Behavior

Emotion regulation in general can be viewed as a component of emotional competence (Dimitrova & Lüdmann, 2014). Thus, emotion regulation is important for a variety of developmental variables, including social functioning (e.g., Eisenberg & Fabes, 1995), mental health (e.g., Kun & Demetrovics, 2010), and stress resilience (e.g., Carlson, Dikecligil, Greenberg, & Mujica-Parodi, 2012). Moreover, emotion regulation inherently influences the occurrence and intensity of experienced emotions (e.g., Gross, 2002), which, in turn, can shape behavior (Lemerise & Arsenio, 2000).

To date, several studies have examined a potential link between emotion regulation and aggressive behavior (for a review, see Roberton, Daffern, & Bucks, 2012). Basically, a more maladaptive or a less adaptive regulation of emotions is often observed in aggressive children, in comparison to their nonaggressive peers. For example, aggressive children focus more on frustrating stimuli and tend to vent their anger (Helmsen & Petermann, 2010); they distract themselves from the frustrating stimuli less often and act less problem-oriented (de Castro et al., 2005). Moreover, several prospective studies emphasize the predictive value of emotion regulation for the development of aggression (for a review, see Röll, Koglin, & Petermann, 2012). In their study, McLaughlin, Hatzenbuehler, Mennin, and Nolen-Hoeksema (2011) examined 11- to 14-year-old children at two time points separated by seven months. Both emotion regulation and aggressive behavior were measured via self-reports. Maladaptive emotion regulation was predictive of subsequent aggressive behavior, whereas aggressive behavior was not predictive of subsequent emotion regulation. In another study, Calvete and Orue (2012) examined 13- to 17-year-old adolescents at two time points separated by six months. Aggressive behavior was assessed by self-reports and peer nominations. Anger regulation was assessed by responses to hypothetical scenarios. Adaptive anger regulation was prospectively linked to less aggressive behavior. In conclusion, emotion regulation in general and anger regulation in particular seem to be negatively correlated with aggressive behavior.

The link between anger regulation and aggressive behavior in childhood can be explained theoretically by applying the prominent SIP model by Crick and Dodge (1994). Lemerise and Arsenio (2000) extended the cognitivistic SIP model by the role of emotion. The authors propose that emotional processes, in particular the regulation of emotions, play an important role in SIP. Accordingly, affective states can influence the SIP steps. A maladaptive regulation of anger often prevents an effective reduction of anger, resulting in maintained or even increased levels of anger (e.g., Denson, Moulds, & Grisham, 2012). This anger can, for example, lead to more hostile attributions of ambiguous social behavior (step 2), the setting of more harmful goals (step 3), the generation of more aggressive responses (step 4), and the evaluation of those aggressive responses as more beneficial (step 5). Thus, a maladaptive in comparison to an adaptive regulation of anger is expected to lead to an increased probability to act aggressively in social situations caused by distorted SIP.

2.2.4 Summary: Emotion and Anger Regulation

Emotion regulation describes processes that influence form, occurrence, intensity, and expression of emotions (Gross, 1998). In past research, specific emotion regulation strategies are distinguished in adaptive and maladaptive strategies depending on the efficiency of the regulatory attempts. A strategy is considered to be maladaptive when it is ineffective in general or in the given situation, when it is socially inappropriate, and when it inhibits functional development (Cole & Hall, 2008). With increasing age, emotion regulation usually gets more adaptive by expanding the repertoire of potential regulation strategies (e.g., Brenner & Salovey, 1997). However, there are stable individual differences in the adaptivity of an individual's emotion regulation (Cole et al., 1994). The emotion anger is one of the emotions that are most difficult to regulate, especially in childhood (e.g., Waters & Thompson, 2014). In addition, the experience of anger is closely related to aggressive behavior (Potegal et al., 2010). Thus, the habitual regulation of anger may be particularly important in the development of aggression. According to the SIP model (Crick & Dodge, 1994), a maladaptive anger regulation should result in biased SIP, which increases the probability to show aggressive behavior. Although there is some empirical evidence for this assumption, there is a lack of assessments of anger regulation that are ecologically valid. Thus, in a first step, the present doctoral dissertation aims to develop a valid assessment of anger regulation strategies for the use in longitudinal studies, followed by the examination of maladaptive anger regulation as a potential risk factor of children's aggressive behavior in a second step.

2.3 Theory of Mind

Understanding that beliefs can differ from reality and, thus, are not explicit representations of the real world, is an important milestone in children's cognitive development (Miller, 2009). The term *theory of mind* refers to that understanding in that it describes the ability of "imput[ing] mental states to [one]self and to others" (Premack & Woodruff, 1978, p. 515). In this sense, theory of mind enables one "to predict what others are going to do on the basis of their desires and beliefs" (Frith & Frith, 2010, p. 165).

To be more specific, theory of mind includes the understanding that different people have different desires, beliefs, and knowledge; that they may hold false beliefs, hide their actual emotions, and sometimes use irony (Peterson, Wellman, & Slaughter, 2012). As a consequence of these complex understandings, children are enabled to show advanced social behavior, including keeping a secret, lying, using irony, coordinating with others, and pretending to be someone else in a role play (e.g., Miller, 2009).

In recent research, theory of mind has often been categorized into two forms, namely cognitive and affective theory of mind (e.g., Shamay-Tsoory, Harari, Aharon-Peretz, & Levkovitz, 2010). The cognitive form refers to inferring others' beliefs and intentions. In contrast, affective theory of mind refers to inferring others' emotional states. Evidence for an independence of these two forms comes primarily from neuroscientific studies showing that, despite a strong overlap, affective and cognitive theory of mind differ in their activation of neural networks (e.g., Kalbe et al., 2010; Sebastian et al., 2012).

2.3.1 Development of Theory of Mind

In past research, the investigation of theory of mind was done using a variety of measures. In general, the subjects were presented a scenario in which the intentions, beliefs, or emotions of one or more characters should be identified. One of the most common measures for cognitive theory of mind is the test for false belief comprehension (for a review, see Miller, 2009). In these scenarios, one person holds a false belief about a specific aspect in the world (first-order false belief), or a false belief about another person that holds a belief about that aspect (second-order false belief). For example, in the prominent Maxi scenario by Wimmer and Perner (1983), the character Maxi puts chocolate in box A. After Maxi left the scene, his mother displaces the chocolate from box A to box B. The first-order false belief comprehension task for the child would be to indicate the box in which Maxi is going to look for the chocolate (box A would be

correct). An alternative way to measure theory of mind is to ask for a prediction of a character's behavior (e.g., Sebastian et al., 2012). This approach has the advantage to assess cognitive and affective theory of mind in a comparable manner in that the prediction can be based on the character's intentions (cognitive) or emotions (affective).

In her review, Slaughter (2015) argues that theory of mind starts to develop in the second year of life as indicated by implicit measures, such as eye-gaze, playing and pointing behavior. For instance, at this age, children often engage in pretend play (e.g., pretending a wooden block to be a toy car), which implies that they understand the difference between thoughts and reality (Bergen, 2002). More advanced theory of mind skills, such as the explicit understanding of first-order false beliefs, usually develops in the age of 3 to 5 years (for a review, see Miller, 2009). Second-order false belief is developing about 1 to 2 years later, in the age of 5 to 6 years. In school age, theory of mind is further developing, and includes the understanding of bluffs, white lies, ambivalent affective states, and the interpretation of moral dilemmas (for a review, see Hughes & Leekam, 2004). In non-clinical populations this course of development seems to be universal (Miller, 2009).

Regarding the differentiation between cognitive and affective theory of mind, the developmental model by Shamay-Tsoory et al. (2010) proposes that functional cognitive theory of mind is a prerequisite for the development of affective theory of mind. Accordingly, inferring emotional states of another person requires the understanding of the beliefs and intention of that other person in the first place. Thus, cognitive compared to affective theory of mind is supposed to develop at an earlier age. In fact, only few studies directly examine this notion. However, there is some support that preschoolers who have success in cognitive theory of mind tasks still have some difficulties in affective tasks (for a review, see Miller, 2009).

Although the general course of theory of mind development seems to be universal in non-clinical populations (Miller, 2009), there are individual differences in theory of mind skills that are relatively stable over time (e.g., Marcovitch et al., 2015). Besides age, the development of theory of mind can be influenced by a variety of other variables, such as the frequency of talks about mental state (Hughes & Dunn, 1998). It is noteworthy that the development of theory of mind does not end in childhood, but is ongoing throughout life (e.g., Henry, Phillips, Ruffman, & Bailey, 2013; Sebastian et al., 2012).

2.3.2 Theory of Mind and Aggressive Behavior

As theory of mind is conceptualized as a core ability in children's socio-cognitive development, there are numerous studies examining potential relations to a broad range of social outcomes. For example, children with lower scores in theory of mind are often found to experience more social difficulties, including higher shyness (e.g., Kokkinos, Kakarani, & Kolovou, 2016), lower peer acceptance (e.g., Caputi, Lecce, Pagnin, & Banerjee, 2012), and lower social competence (e.g., Devine, White, Ensor, & Hughes, 2016).

With regard to aggressive behavior, there are already many studies investigating the link between theory of mind and aggression. However, the results are fairly mixed. In the research on bullying, a form of aggressive behavior characterized by power imbalance and repetitiveness (Olweus, 2013), there is some evidence for a positive association between the two constructs. Using peer nominations, Sutton, Smith, and Swettenham (1999a) categorized their 7- to 10-year-old participants in a specific bullying role, including bully leaders, bully assistants, and bully victims. What they found was that bully leaders – that is, children who take initiative in bully behavior – were characterized by better theory of mind skills than the assistants and victims. This result is compatible with the idea that superior theory of mind can enable children to manipulate others more effectively (Sutton, Smith, & Swettenham, 1999b). In particular, this reasoning especially applies to relational forms of aggression as this form requires the involvement of other peers (e.g., spreading rumors requires not only the perpetrator and victim, but also a third party). This notion found support in studies by Renouf et al. (2010) who found theory of mind being positively correlated with relational, but not physical aggression in kindergarten children.

In contrast to the studies mentioned above, there are some studies that document a negative association between theory of mind and aggressive behavior. A recent study by O'Toole, Monks, and Tsermentseli (2017) found a negative link between theory of mind and physical aggression in early childhood; for relational aggression, no significant link was found. However, Kokkinos, Voulgaridou, Mandrali, and Parousidou (2016) specifically examined the association between theory of mind and relational aggression in pre-adolescents and found a negative correlation.

The SIP model by Crick and Dodge (1994) offers a theoretical explanation for a negative link between theory of mind and aggressive behavior in childhood. High scores in theory of mind equals a good understanding of the beliefs, intentions, and emotions of others. This understanding can reduce biases in SIP, in particular in the early steps of the SIP model. More

specifically, the interpretation of social cues (SIP step 2) can be biased if an individual has problems with inferring the intentions of others. For example, a child that is accidentally pushed by a peer may incorrectly interpret this behavior as a hostile provocation. Choe, Lane, Grabell, and Olson (2013) supported this notion by demonstrating that preschoolers with inferior theory of mind skills are more likely to make such hostile attributions. In addition, theory of mind can play a role when it comes to deciding for a response (SIP step 5). For instance, anticipating the resulting emotions of others and considering their beliefs about one's own behavior may help in deciding for an appropriate response. This reasoning is supported by Grueneisen, Wyman, and Tomasello (2015), who could show that 6-year-old children adjust their response decisions in a coordination game by considering false beliefs of their play partner.

In conclusion, according to the SIP model, low levels of theory of mind are viewed as a social deficit that biases SIP. That, in turn, may foster the development of aggressive behavior. However, considering the reasoning by Sutton et al. (1999b), the association between theory of mind and aggressive behavior may differ depending on the specific form of aggressive behavior. Relational aggression usually requires the involvement of others, for example when it comes to spreading rumors. The opportunities to show efficient relational aggression are increased with superior levels of theory of mind, that is, with a better understanding of others' beliefs and emotions. In contrast, physical aggression does not require the involvement of third parties and thus, superior theory of mind skills should not increase the efficiency of this aggression form.

2.3.3 Summary: Theory of Mind

Theory of mind describes the ability to attribute mental states to the self and to others; that is, to understand that different people have desires, beliefs, intentions, and emotions that may differ from the self and the real world (Peterson et al., 2012). In this sense, theory of mind is viewed as a core milestone in children's cognitive development as it enables them to show advanced social behavior, such as coordinating with others (Miller, 2009). In the course of early childhood, a fundamental understanding of others' mental states develops (Miller, 2009). A broad range of research emphasizes the importance of theory of mind for the development of social adjustment. Regarding aggressive behavior, however, empirical results are mixed. According to the SIP model (Crick & Dodge, 1994), problems in theory of mind should result in biased SIP, which increases the probability to show aggressive behavior. However, in examining this relation the differentiation in forms of aggression may be of considerable

relevance. Specifically, superior theory of mind may be advantageous for manipulating others effectively and, thus, may be associated with increased relationally aggressive behavior (Sutton et al., 1999b). To date, there is no study that has investigated potential reciprocal relations between theory of mind and the engagement in different forms of aggression over the period of middle childhood. Therefore, the present doctoral dissertation aims to examine differentiated associations between these two constructs using a longitudinal study design.

2.4 Physical Attractiveness

Physical attractiveness – in the following also abbreviated as attractiveness – describes "the extent to which a stimulus person is pleasing to observe" (Patzer, 2006, p. 324). This short definition stresses three important aspects of the construct. First, attractiveness is assumed to trigger positive affective states via the visual sensory channel ("pleasing to observe"). Thus, attractiveness is supposed to be inherently positive. Second, for scientific use, the term attractiveness refers to the description of human beings ("stimulus person") rather than inanimate objects or non-human species. Third, attractiveness is viewed as a continuous variable ("extent"), not as a sheer dichotomous construct. That is, we do not only distinguish between unattractive and attractive. Instead, we evaluate the individual level of attractiveness based on a continuum from low to high.

As physical attractiveness is primarily perceived through the visual sensory channel, a person's configuration of physical (i.e., bodily and facial) characteristics are central in the perception of his or her attractiveness (e.g., Mehrabian & Blum, 1997). Thus, there are multiple studies investigating physical determinants of attractiveness. A basic condition for such examinations is the existence of social consensus regarding the evaluation of another individual as less versus more attractive. Indeed, early studies revealed that there is a considerably high agreement between different assessors ("truth-of-consensus", for a review, see Patzer, 1994). The existence of a social consensus in attractiveness evaluation is also reflected in the fact that computer software is even able to give a fairly accurate estimation of an individual's physical attractiveness (e.g., Eisenthal, Dror, & Ruppin, 2006). Agreement on attractiveness evaluations is also observable across age groups. For example, Slater et al. (1998) found that newborns as young as a few days already show a preference for faces that are evaluated as attractive by adults in comparison to those that are evaluated as unattractive.

The vast majority of studies examining physical or situational determinants of attractiveness are aimed at adult populations (for a review for adult attractiveness, see Agthe, Aydin, Pfundmair, Frey, & Niesta Kayser, 2016; Perrett, 2012; Swami & Furnham, 2008). Only a few studies dealing with an earlier age group. However, in general, faces of infants and children are considered to be more attractive if they are more compatible with the baby schema (i.e., large eyes and forehead, and small nose and mouth; e.g., Glocker et al., 2009; Hildebrandt & Fitzgerald, 1979). This is true for both girls and boys. However, with age, the positive relation between babyishness and perceived attractiveness seems to increase for the evaluation of girls', whereas it decreases for the evaluation of boys' attractiveness (Rieger et al., 2011). Besides the question what is considered to be attractive and what is not, social psychology has documented the social impact of physical attractiveness quite intensively, as described in the following.

2.4.1 The Impact of Physical Attractiveness

Physical attractiveness is considered to be a very salient and quickly perceived characteristic. Olson and Marshuetz (2005) argue based on their studies that attractiveness can hardly be ignored since, even under worst viewing conditions (e.g., by masking stimulus material) and with only a few milliseconds of presentation time, participants were able to correctly assess the stimulus' attractiveness. Keeping that in mind, it does not seem surprising that attractiveness can have a considerable impact on social impression formation.

In their prominent study, Dion, Berscheid, and Walster (1972) could demonstrate the existence of a *physical attractiveness stereotype*. That is, the perception of less versus more attractive individuals is linked to specific expectations about these individuals. Basically, high levels of physical attractiveness usually trigger more positive expectations than low levels of attractiveness ("What is beautiful is good"). A meta-analysis by Eagly, Ashmore, Makhijani, and Longo (1991) revealed that the biggest effect of attractiveness is on the perception of social competence, followed by potency, adjustment, and intellectual competence. In contrast, the perception of integrity and concerns for others are only little influenced by attractiveness. The specific content of the stereotype, however, can vary with culture (e.g., Wheeler & Kim, 1997). The observation that more attractive individuals are perceived as more positive is consistent with study results from neuroscience and cognitive science. Accordingly, the perception of an attractive face can activate the reward system in the brain (for a review, see Hahn & Perrett, 2014) and function as a reliable prime for positive valenced words (e.g., Olson & Marshuetz, 2005).

In addition to the positive social effects of high physical attractiveness, low attractiveness is not only associated with the absence of positive characteristics, but also with the presence of negative ones ("What is ugly is bad"). For example, Dion (1973) found that 3- to 6-year-old children attribute aggressive behavior more often to unattractive than attractive unfamiliar peers. Griffin and Langlois (2006) concluded that the negative effects for unattractiveness may be greater than the positive effects of high attractiveness. In one of their experimental studies, 7- to 9-year-old children evaluated unattractive faces as less sociable, altruistic, and intelligent in comparison to medium attractive faces. In contrast, attractive faces were only viewed as more sociable than medium faces, but there were no effects for perceived altruism and intelligence. This negative stereotype is consistent with a recent study by Schein and Langlois (2015) demonstrating that the presentation of unattractive infant faces can trigger facial muscle movements in adults that are linked to negative affect.

The impact of physical attractiveness is not limited to sheer perception and generation of expectations. Numerous studies found convincing evidence that an individual's attractiveness also influences the treatment the respective individual receives by others. A meta-analysis by Langlois et al. (2000) concluded that both more attractive adults and children relative to their less attractive counterparts are in general treated more favorably by their environment. For example, attractive children receive more lenient sanctions for misconducts (e.g., Berkowitz & Frodi, 1979), preferential treatment by their teachers (e.g., Dunkake, Kiechle, Klein, & Rosar, 2012), and even experience more affectionate caregiving by their parents (Langlois, Ritter, Casey, & Sawin, 1995).

It seems reasonable to assume that these differences in treatment by others can influence the individual development of a person. Benzeval, Green, and Macintyre (2013) assessed the physical attractiveness of 15-year-old adolescents by adult ratings. About 20 years later, they assessed several life outcomes. The authors found that the level of attractiveness in adolescence was a positive predictor of most of the outcomes in later adulthood, including marital status, educational, and occupational success. Thus, attractive individuals in contrast to their unattractive counterparts develop more favorably in many areas of life, with the most conclusive empirical support existing for the social domain. In their meta-analysis, Langlois et al. (2000) found the largest effect size ($d \sim .70$) for the association between attractiveness and social advantages, in both children and adults. Accordingly, more attractive compared to less attractive individuals are more popular among their peers, have more friends, and a higher social

standing in their peer group. In contrast, less attractive children are at a higher risk for experiencing peer problems (e.g., Sweeting & West, 2010).

2.4.2 Physical Attractiveness and Aggressive Behavior

So far, only few studies investigated a potential correlation between an individual's level of physical attractiveness and its involvement in aggressive behavior. In a study by Borch, Hyde, and Cillessen (2011), a positive correlation between the physical attractiveness of adult students and their exhibition of aggressive behavior was found. However, Bobadilla, Metze, and Taylor (2013) found a negative correlation for adult students, but only in males. Studies regarding childhood also revealed a negative association between attractiveness and aggression. For example, Langlois and Downs (1979) observed 3- and 5-year-old children during dyadic free play. They coded several behaviors, including aggressive behavior. The authors found no relation between attractiveness and aggression in the 3-year-olds. However, in the 5-year-olds they found that unattractive children act more aggressively towards their play partner than attractive children. This result was more pronounced in boys than in girls. Similarly, another set of studies showed that children, who were referred for their aggressive behavior by their teachers, received lower scores on physical attractiveness by independent student raters than the nonaggressive children (Dumas, Nilsen, & Lynch, 2001; Serketich & Dumas, 1997).

There are a number of theories that may explain a possible association between attractiveness and aggression (for a review, see Langlois et al., 2000). Evolutionary theories usually claim that physical attractiveness and specific personality traits or behavioral patterns inherently co-occur in an individual, for example caused by common genes or mate selection processes (e.g., Kanazawa & Kovar, 2004). In contrast, socialization theories focus on social processes that are influenced by an individual's level of attractiveness, and in turn shape the development of that individual. In this context, self-fulfilling prophecies may function as the core of the socialization process (e.g., Darley & Fazio, 1980; for a schematic overview, see Figure 2). In a social situation, the level of physical attractiveness of an individual (target) activates the partner's attractiveness stereotype. That is, if the target is viewed as attractive, the partner will generate mainly positive expectations. However, if the target is viewed as unattractive, the partner will generate more negative expectations. These expectations, in turn, will bias the perception and, in consequence, the treatment the target will receive from the partner. In addition, the motivation to get in touch and invest in building a social relationship with the target depends on these expectations. As a result, the target gets further opportunities

to develop stereotypically. That is, for less attractive individuals the activation of the "What is ugly is bad" stereotype leads to negative expectations and, consequently, to unfavorable treatment. In contrast, for more attractive individuals the activation of the "What is beautiful is good" stereotype leads to positive expectations and, as a consequence, to more favorable treatment. As a result, the less compared to the more attractive individual is more exposed to a social environment that may foster the development of aggressive behavior. In this sense, socialization theories would predict a negative association between attractiveness and aggression.

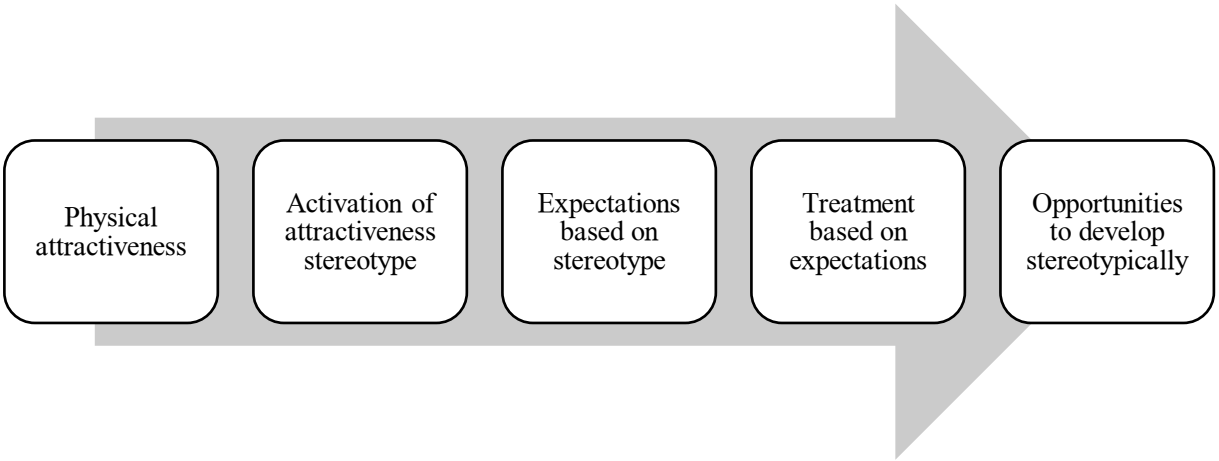

Figure 2. Schematic overview of the self-fulfilling prophecy of physical attractiveness

This reasoning is fully compatible with the SIP model by Crick and Dodge (1994). The probability of experiencing problematic social situations is higher for less attractive as compared to more attractive individuals (e.g., Sweeting & West, 2010). The memories of these social situations and experiences are stored in the database and influence all SIP steps. That is, in comparison to more attractive children, the database of less attractive children may contain more memories of unfavorable social experiences. This, in turn, may lead to biased SIP. For example, unattractive in contrast to attractive children may interpret ambiguous social behavior of others in a more hostile way (SIP step 2), they may generate fewer socially competent behaviors for the social situation (SIP step 4), and they may miss the decision for actually beneficial response decisions caused by worse outcome expectations (SIP step 5). In addition, the quality of the social relationship influences the goal setting (SIP step 3). As less attractive children experience more problems with their peers (e.g., Sweeting & West, 2010), they may tend to generate goals that are not primarily aimed to improve the relationship.

In this sense, the SIP model can be fully implemented in the superordinate socialization processes described above. After perceiving differential treatment based on one's own level of attractiveness, SIP is biased in more disadvantaged directions, which increases the probability to show aggressive behavior.

2.4.3 Summary: Physical Attractiveness

An individual's physical attractiveness is an appearance-related characteristic that describes the extent of pleasure that is triggered by observing the respective individual (Patzer, 2006). Early research revealed the existence of an attractiveness stereotype (Dion et al., 1972). That is, when observing individuals, specific cognitions and expectations about them emerge based on their level of attractiveness. This stereotype also manifests in differential treatment by others. To date, there are only few studies examining a correlation with aggressive behavior, and these yielded mixed findings. According to socialization theories (e.g., Langlois et al., 2000), a negative association between attractiveness and aggression is expected that can be explained by self-fulfilling prophecies. Physical attractiveness activates the attractiveness stereotype, resulting in corresponding expectations and, consequently, in differentiated social treatment. The resulting social experiences are stored in the database, as assumed by the SIP model (Crick & Dodge, 1994). Since the database strongly affects SIP, SIP may be biased. In consequence, the probability of aggressive behavior is influenced. Thus, more attractive children that trigger favorable expectations and treatment by others are supposed to develop less aggressively. In contrast, less attractive children trigger unfavorable expectations and treatment by others, which in turn may lead to a more aggressive development. This notion was not yet tested in any previous study. Hence, the present doctoral dissertation aims to examine physical attractiveness as a distal developmental factor of aggressive behavior by taking problematic peer interactions into account. In this sense, physical attractiveness is assumed to be an intrapersonal characteristic that unfold its impact on aggression through interpersonal processes.

3

Aims of the Doctoral Dissertation

The primary aim of this doctoral dissertation is to identify potential risk factors of aggressive behavior in childhood. Previous research already examined numerous risk factors of aggression, including intra- and interpersonal as well as environmental factors (for a review, see Krahé, 2013; Valois, MacDonald, Bretous, Fischer, & Drane, 2002). As a theoretical framework to predict aggressive behavior the prominent SIP model by Crick and Dodge (1994) is used. In terms of this model, intrapersonal characteristics seem to be considerably influential on SIP. Thus, the present work focused on intrapersonal factors, that is, factors that are inherently attached to the individual.

According to the SIP model, there are six steps in a given social situation that connect the occurrence of a social cue to the resulting social behavior: encoding and interpretation of social cues, clarification of goals, response access or construction, response decision, and behavioral enactment (see Figure 1 in Chapter 2.1.3). Both cognitive, affective, and appearance-related individual differences can have an impact on these SIP steps and, in consequence, may be important to understand the occurrence and development of aggression. This work aims to consider all three domains by examining one particular construct of each domain. For the affective domain children's *anger regulation*, for the cognitive domain children's *theory of mind*, and for the appearance-related domain children's *physical attractiveness* are examined as potential developmental factors of aggression. These constructs were chosen because they function as prototypical representatives of the three superordinate domains, as described in the following subchapters. Each empirical paper of this doctoral dissertation is focused on one of the three constructs.

In addition to the examination of potential risk factors, another major goal of this work is the development and validation of an age-appropriate alternative to the behavioral assessment of anger regulation – originally presented by Rohlf and Krahé (2015) – for the use in longitudinal studies. Thus, in the following, the specific aims for the development of this measure are introduced (Paper 1), followed by the ideas and aims regarding the identification of risk factors of aggressive behavior (Paper 1-3). Table 1 presents an integrative overview of the specific hypotheses that are proposed for the empirical papers.

Table 1. Overview of proposed hypotheses

Paper 1	
Hypothesis 1.1	Equivalence of the two behavioral observation assessments of anger regulation (metric invariance)
Hypothesis 1.2	Substantial correlation between the two measures over three years (construct stability)
Hypothesis 1.3	Substantial correlations between anger regulation and aggressive behavior, peer problems, and conduct problems (criterion validity)
Hypothesis 1.4*	Maladaptive anger regulation predictive of aggressive behavior (+)
Paper 2	
Hypothesis 2.1	Theory of mind predictive of physical aggression (-)
Hypothesis 2.2	Theory of mind predictive of relational aggression (-/+)
Hypothesis 2.3	Aggressive behavior predictive of theory of mind (-)
Paper 3	
Hypothesis 3.1	Physical attractiveness predictive of peer problems (-)
Hypothesis 3.2	Peer problems predictive of aggressive behavior (+)
Hypothesis 3.3	Indirect effect from attractiveness to aggression via peer problems (-)

Notes. * not included in the empirical paper; (-) = a negative relation is hypothesized, (+) = a positive relation is hypothesized, (-/+) = no specific direction of relation is hypothesized.

3.1 Development of an Anger Regulation Measure for the Use in Longitudinal Studies

So far, there is a lack of ecologically valid measures to assess anger regulation as a potential developmental factor of aggression. Previous studies predominantly used self-reports (Aldao et al., 2010). However, such measures are fraught with methodological problems, especially in the examination of children (Underwood, 1997; see Chapter 2.2.2). That is why Rohlf and Krahé (2015) sought to develop a behavioral observation assessment for children's use of specific anger regulation strategies. The children were confronted with an anger-eliciting situation – more specifically, a virtually impossible dexterity task – and their strategies to deal with the anger were coded. This behavioral observation assessment was found to have convincing psychometric properties in 6- to 10-year-old children (reliability, construct and criterion validity; Rohlf & Krahé, 2015; Kirsch et al., 2015). The sample that was examined in these studies is the same that is used in the present study on anger regulation (Paper 1).

The present doctoral dissertation aims to identify intrapersonal risk factors using a longitudinal study design. To control for potential reciprocal effects that may occur over time (Selig & Little, 2012), it is important to repeat the anger regulation assessment rather than assessing it only at the first measurement point. However, for the use in longitudinal studies, it is inherently problematic to present the exact same task repeatedly due to the danger of increased reactivity that would diminish internal validity (Schaie & Hofer, 2001). Furthermore, when the children get older, they become more cognitively and motorically skilled (e.g., Bartolotta & Shulman, 2010), challenging the age appropriateness of tasks that are designed for younger children. As a result, the development of a sufficiently different, but conceptually equivalent anger-eliciting task is important for examining developmental courses.

In line with this reasoning, the major goal of Paper 1 is to further develop the behavioral observation method by Rohlf and Krahé (2015) for the use in longitudinal studies. That is, the previously developed assessment is supposed to be adapted and validated for children of an older age. At the same time, both assessments are required to be conceptually equivalent, and to be substantially correlated with one another as emotion regulation is conceptualized as a relatively stable individual difference variable (Cole et al., 1994). Thus, in Paper 1 the following hypotheses are examined (see Table 1): (1) Both assessments of anger regulation are conceptually equivalent to one another. On a statistical level, at least metric measurement invariance is aimed to be achieved because this level of invariance is required to compare regression slopes longitudinally (Chen, 2007). (2) Both assessments of anger regulation are substantially correlated, indicating construct stability over time. (3) There are substantial

concurrent and prospective correlations between anger regulation and constructs that are theoretically and empirically assumed to be associated with anger regulation (criterion validity). In particular, the correlations with aggressive behavior, peer problems, and conduct problems are considered.

3.2 Identification of Risk Factors of Aggressive Behavior

3.2.1 Anger Regulation as an Affective Factor

Based on the developed assessment described above, maladaptive anger regulation was examined as a risk factor of aggressive behavior in the affective domain. Within the last three decades, emotion regulation received an enormous increase in scientific attention (Adrian et al., 2011). In this context, emotion regulation in general was frequently investigated as a correlate of the occurrence and development of aggressive behavior (for a review, see Röll et al., 2012). Regarding the specific emotion of anger, only a few studies exist so far. However, among all emotions, anger is considered to be most closely associated with aggression (e.g., Stemmler, 2010). Therefore, it seems instructive to explicitly examine the regulation of anger as a potential risk factor of aggression.

According to the SIP model (Crick & Dodge, 1994) and its extended version (Lemerise & Arsenio, 2000), emotional processes can have a considerable impact on SIP, and, thus, can play an important role for the development of social behavior (see Chapter 2.2.3). For example, anger in a social situation can lead to increased hostile attribution biases, which in turn are associated with the occurrence of aggressive behavior (for a meta-analytic review, see de Castro et al., 2002). If children are not able to regulate their angry feelings effectively (i.e., they show maladaptive anger regulation), the anger will persist or even increase (e.g., Denson et al., 2012). Thus, anger regulation can be assumed to predict the possibility of aggressive behavior by leading to biased SIP.

In sum, the research investigating the link between emotion regulation and aggressive behavior in middle childhood used primarily self-reports and – for the most part – examined emotion regulation in general rather than anger regulation in particular. In addition, most research is cross-sectional. Thus, the prospective (i.e., longitudinal) examination of the specific association of anger regulation and aggression by using an ecologically valid measure is of scientific relevance.

The research question whether anger regulation is predictive of the development of aggression is only marginally considered in Paper 1 as part of the validation procedure of the behavioral observation measures. A more elaborate analysis is not in the scope of this methodological paper. However, Appendix A provides further analyses that examine the following hypothesis: Maladaptive anger regulation is a positive predictor of subsequent aggressive behavior. That is, children who face problems with the adaptive regulation of their angry feelings are at a higher risk to behave aggressively at a later occasion.

3.2.2 Theory of Mind as a Cognitive Factor

For the domain of cognitive risk factors of aggressive behavior, deficits in theory of mind were examined. The development of this socio-cognitive construct is viewed as essential as it represents the fundamental ability to attribute mental states to the self and to others. As a consequence of this understanding, children are enabled to show advanced social behavior, including coordinating with others, keeping a secret, or lying (e.g., Miller, 2009). Despite a normative developmental course (Miller, 2009), there are individual differences in theory of mind that are relatively stable over time (e.g., Marcovitch et al., 2015). Multiple studies found that deficits in theory of mind are associated with a broad range of negative developmental outcomes, such as low social competence (e.g., Devine et al., 2016). In this context, the development of aggressive behavior also may be influenced by theory of mind.

According to the SIP model (Crick & Dodge, 1994), theory of mind may be of considerable relevance for the emergence of aggressive behavior (see Chapter 2.3.2). A lack of understanding of others' beliefs, intentions, and emotions can bias SIP in a way that increases the possibility of aggressive responses. For example, children that have difficulties to infer the intentions of others may generate an incorrect, possibly more hostile interpretation of the behavior of that other person (e.g., Choe et al., 2013). In addition, deficits in theory of mind may also result in problems in considering others' beliefs when it comes to decide on an appropriate social response (e.g., Grueneisen et al., 2015).

This reasoning suggests that deficits in theory of mind would result in biased SIP that, in turn, would lead to more aggressive behavior in general. However, in their criticism, Sutton et al. (1999b) argue that superior theory of mind could be advantageous to manipulate others effectively and, thus, may fuel relationally aggressive behavior. On an empirical level, there is mixed evidence for these two competing ideas. Whereas physically aggressive behavior is

usually found to be negatively correlated with theory of mind (e.g., O'Toole et al., 2017), the findings for the correlation with relationally aggressive behavior are somewhat inconclusive. Some studies found positive (e.g., Renouf et al., 2010) and some negative (e.g., Kokkinos et al., 2016) associations. In conclusion, it seems that different forms of aggression may be differently associated with theory of mind skills.

Even though the direction of effects is still in question, this discussion implies theory of mind to be a predictor of aggressive behavior. However, in addition to this notion, it may also be possible that the involvement in aggressive behavior is predictive of subsequent theory of mind skills. In their conceptualization, Crick and Dodge (1994) emphasize that social experiences, as they may arise from the involvement of aggression, shape the database and individual differences in SIP, respectively. Theory of mind is a core cognitive ability that is involved in most of the SIP steps and, therefore, its development may be influenced by the engagement in aggressive behavior. This notion found empirical support in a study by Lansford et al. (2010), showing that aggressive behavior and components of SIP influence one another over the course of time.

In line with this reasoning, Paper 2 aimed to examine reciprocal relations between theory of mind and aggressive behavior in childhood by distinguishing two forms of aggressive behavior, namely physical and relational aggression. Specifically, in Paper 2 the following hypotheses are examined (see Table 1): (1) Theory of mind is a negative predictor of subsequent physical aggression in childhood. (2) Theory of mind is a predictor of subsequent relational aggression in childhood. In contrast to physical aggression, theoretical and empirical considerations suggest both a positive and negative association with relational aggression. (3) Aggressive behavior (both physical and relational) is a negative predictor of subsequent theory of mind.

3.2.3 Physical Attractiveness as an Appearance-Related Factor

For the domain of appearance-related risk factors of aggressive behavior, a lack of physical attractiveness was examined. Attractiveness is considered to be an important aspect of physical appearance as it describes the extent to which an individual is pleasing to observe (Patzer, 2006). Multiple studies emphasize the importance of physical attractiveness in the development of social relationships, even in childhood (e.g., Sweeting & West, 2010). Social relationships, in turn, can have a remarkable impact on the development of an individual's health and

adjustment (e.g., Umberson & Montez, 2010). Thus, attractiveness may be a distal developmental factor for the emergence and maintenance of subsequent maladjustments.

This causal chain found support in a study by Rosen, Underwood, and Beron (2011), in which school-aged children were rated for attractiveness by independent adult raters. Teachers assessed the children's degree of internalizing problems and their experience of victimization. Less attractive in comparison to more attractive children were rated as experiencing more victimization, which in turn led to more internalizing problems. Problematic peer relationships, however, are not only linked to the development of internalizing behavior, but also to the development of externalizing behavior (for a review, see Leary, Twenge, & Quinlivan, 2006).

This reasoning is consistent with socialization theories (e.g., Langlois et al., 2000; see Chapter 2.4.2). Accordingly, the level of attractiveness may be predictive of developmental problems. The perception of physical attractiveness activates the attractiveness stereotype, which in turn lead to differential expectations and cognitions. That is, an attractive child is expected to have more positive traits, whereas an unattractive child is expected to have more negative ones. These appearance-dependent expectations lead to corresponding treatments. While the attractive child is treated more favorably, the unattractive child is treated more unfavorably by his or her social environment. Consequently, children have different social experiences based on their appearance that they store in their memory. In terms of the SIP model by Crick and Dodge (1994), these memories are part of the database and influence subsequent SIP. For example, in one of their studies, Dodge et al. (2003) could demonstrate that low social preference in middle childhood predicts subsequent problematic SIP patterns (i.e., hostile attribution bias, generation and enactment of less competent responses). Problematic SIP patterns, in turn, predicted aggressive behavior.

In line with this reasoning, the aim of Paper 3 was to examine physical attractiveness as a distal appearance-related factor for the development of aggressive behavior. In this context, the experience of peer problems is considered as an underlying mechanism, linking attractiveness to aggression. Specifically, Paper 3 examined the following hypotheses (see Table 1): (1) Physical attractiveness is a negative predictor of the experience of subsequent peer problems in childhood. That is, less attractive relative to more attractive children are at a higher risk to experience problems with their peers. (2) The experience of peer problems is a positive predictor of subsequent aggression in childhood. That is, children who experience more problems with their peers are at a higher risk to behave aggressively. (3) There is a negative indirect effect of physical attractiveness on the development of aggression through experienced

peer problems. In contrast to more attractive children, less attractive ones behave more aggressively as a result of their more problematic peer relationships.

4

Research Concept

The empirical examination of the proposed hypotheses was done using data from the "Potsdamer Intrapersonale Entwicklungsrisiken"-Study (PIER-Project). The PIER-Project was conducted by a graduate college that was funded by the German Research Foundation (grant number GRK 1668). Within the framework of the project, different fields of psychology worked together, including social, developmental, and health psychology. The aim of the PIER-Project was to identify intrapersonal risk factors; that is, risk factors that are related to the individual rather than to his or her environment. In this context, the present doctoral dissertation is focused on identifying intrapersonal risk factors for the development of aggressive behavior.

The PIER-Project contains two central projects. Central Project 1 examined children and adolescents over a period of 11 years (from 2005 to 2016), covering five measurement points. Central Project 2, on the other hand, examined children over a period of 3 years (from 2012 to 2015), covering three measurement points. This doctoral dissertation used data from Central Project 2. Thus, in the following, the shortened term "PIER-Project" is used to refer to the Central Project 2 of the PIER-Project.

In the following subchapters, the sample and procedure of the PIER-Project is presented in more detail, as well as the specific measurement instruments and statistical analyses that were used to examine the hypotheses.

4.1 Sample

At the first measurement point (T1), the total sample of the PIER-Project consisted of 1,657 children. The children were recruited from 33 community elementary schools of the federal state of Brandenburg, Germany. The sample was distributed to 178 different school classes covering first to fourth grade. The mean age of the children was M = 8.36 years (SD = 0.95), with a range from 6 to 11 years. The sex ratio was balanced with a proportion of girls of 52.1%.

At the second measurement point (T2), 1,612 children participated again, which equals an attrition rate of 2.8%. The interval between T1 and T2 was about 9 months (M = 8.52 months, SD = 1.87). At the third and final measurement point (T3), 1,501 children participated again. Thus, the attrition rate from T1 to T3 was 9.5%. The interval between T2 and T3 was about two years (M = 23.15 months, SD = 1.73). Table 2 presents an overview of the central demographic characteristics of the sample for all three measurement points.

Table 2. Description of the total sample of the PIER-Project

	T1	T2	T3
Sample size *N*	1,657	1,612	1,501
Proportion girls	52.1 %	51.9 %	51.8 %
Year of collection	2012	2013	2015
Children's grade	1-4	1-5	3-6
Age range	6-11	7-11	9-13
Age *M* (*SD*)	8.36 (0.95)	9.11 (0.93)	11.02 (0.92)

The data of the total sample were used only in Paper 2. Since in Paper 1 and Paper 3 data were used that included visual material of the children, the total sample could not be used. The primary reason was that not all parents gave permission to record visual material of the children (n = 474 at T1). In addition, the quality of the visual material was not sufficient for all recordings (n = 140 at T1). Thus, the sample was reduced to n = 1,043 at T1. Paper 3 used data from this reduced sample. The core of Paper 1 were two behavioral observations that required time-consuming video coding. Due to limited resources, a randomly chosen subsample at T1 of n = 599 children were included in the video coding procedure (for more details, see Rohlf & Krahé, 2015). In summary, Paper 1 used a reduced sample of N = 599 children, Paper 2 the total sample of N = 1,657 children, and Paper 3 another reduced sample of N = 1,043 children. Paper 1 only included data from T1 and T3, whereas Paper 2 and Paper 3 included data from all three measurement points.

4.2 Procedure

The PIER-Project used a multi-informant approach to obtain data of the children. That is, not only one single source, but three different sources were used. Specifically, data were collected from the children themselves, their parents, and the children's teachers.

Data from the children were collected at their schools. The data collection always took place in isolated, private rooms that were provided by the schools. During data collection, only two persons were in the room, namely the child and the examinator. In total, there were 42 to 59 different examinators depending on the measurement point. Before the data collection started at all three measurement points, the examinators were trained in a standardized examination procedure. Additionally, they were instructed to cultivate a friendly, but neutral

communication with the children. For the self-reports at T1 and T2, children were given the response scale and they respond by pointing at the corresponding response option. The examinators protocolled their response. At T3, the data collection was computer-supported, and the children gave their responses in private. For each questionnaire and test, the children were instructed in a standardized manner by the examinator. During the examination, the children were given little rewards to show gratitude and to keep them motivated. At the end of the examination, the children were given a cinema voucher (T1 and T2) or a local book store voucher (T3).

Data from the parents were collected by sending them questionnaires by mail. They were asked to respond to questions with regard to their child. If parents had more than one child that participated in the PIER-Project, they received independent questionnaires for each child. The parents could fill out the questionnaires by hand or online.

Data from teachers were also collected by sending them questionnaires. For each child that participated in the PIER-Project they received an independent questionnaire. For each completed questionnaire, the corresponding school class received 5€ for its class fund.

The procedure and used instruments were approved by the Ethics Committee of the University of Potsdam and the Ministry of Education, Youth and Sport of the Federal State of Brandenburg. Before each measurement point, the parents of each participating child were informed about the data collection and gave permission to examine the child.

4.3 Overview of Central Instruments

In the following, a concise overview of the instruments that were used to test the hypotheses is presented. First, the instruments for assessing the dependent variable of aggressive behavior are introduced. Second, the instruments for assessing the risk factors of aggression are presented; that is, anger regulation, theory of mind, and physical attractiveness. Finally, as peer problems was a central construct in two papers (for validation in Paper 1 and for mediation in Paper 3, respectively) this construct is also introduced in this subchapter. A more detailed description of each instrument and the description of additional instruments can be retrieved from the empirical papers (see Chapter 5). Table 3 presents summarized information of the central instruments and their assignment to the empirical papers. Verbatim instructions and items can be retrieved from Appendix C.

Table 3. Overview of central instruments used in the empirical papers

	Source	Instrument	Number of items	Response scale	Assignment of measurement points		
					Paper 1	Paper 2	Paper 3
Aggressive behavior	T	CSBS-T	6[a]	1-5	T1, T3	T1, T2, T3	T1, T2, T3
	C	CSBS-T	6[a]	1-4	T3	-	-
	C	In situ	1	0-5	T3	-	-
Anger regulation	C	Observation	5	-	T1, T3	-	-
Theory of mind	C	Cartoons	12[b]	0-1	-	T1, T2	-
Physical attractiveness	R	Rating	1[c]	1-7	-	-	T1, T3
Peer problems	T	SDQ	3	1-3	T1, T3	-	-
	C	FEESS	$8_{(T1/T2)}/7_{(T3)}$	$1\text{-}2_{(T1/T2)}/1\text{-}4_{(T3)}$	T1, T3	-	T1, T2, T3
	P	SDQ	5	1-3	T1, T3	-	-

Notes. Source: T = teacher, C = child, R = independent rater, P = parents; [a] two subscales with 3 items each, [b] two subscales with 6 items each, [c] 24 (T1) and 23 (T3) rater.

4.3.1 Instruments for Aggressive Behavior

Aggressive behavior is the central construct in this doctoral dissertation, and, therefore, included in each empirical paper. To assess children's aggressive behavior, three different measures were used, namely a teacher-report, children's self-report, and a behavioral measure. The teacher-report was used at all three measurement points, whereas the child-report was assessed only at T2 and T3. The behavioral measure was assessed only at T3.

The teacher-report included six items adapted from the Children´s Social Behavior Scale – Teacher Form (CSBS-T; Crick, 1996) and a 5-point response scale ranging from 1 (*never*) to 5 (*daily*). This measure offers a total score for aggressive behavior, but also can be decomposed into two subscales, namely physical and relational aggression. In Paper 1, only the total score was used due to the criterion validation procedure of the newly developed behavioral observation assessment of anger regulation (see Appendix A for a differentiation in the two subscores). Paper 2 used the subscores instead of the total score because it was hypothesized that theory of mind may be associated differently with the two forms of aggressive behavior. In Paper 3, the total score was used to test the hypotheses, but in further exploratory analyses the measure was decomposed into the two subscores.

The child-report essentially included the same six items as the teacher-report, but they were reformulated to be appropriate for the younger recipients. In addition, the response scale was simplified into a 4-point response scale ranging from 1 (*never or almost never*) to 4 (*almost daily*). The child-report was only used in Paper 1 to validate the behavioral observation measure of anger regulation. Since aggressive behavior is obviously a socially undesirable behavior, self-reports can lead to an underreporting of aggression due to social desirability responding (e.g., Vigil-Colet, Ruiz-Pamies, Anguiano-Carrasco, & Lorenzo-Seva, 2012). This may be especially problematic as at T1 and T2, the children gave their responses not privately, as described above. Furthermore, this measure was not used at all measurement points. Due to these two limitations of children's self-report, it was not appropriate to include this measure in the longitudinal analyses that were in the focus of Paper 2 and Paper 3.

The behavioral measure was based on the Tangram Hurt/Help Task by Saleem, Anderson, and Barlett (2015). The children were given a virtually impossible dexterity task (stacking dice to form a stable tower). After they failed in this task, the children were given the opportunity to determine the task's difficulty level for another alleged child (i.e., determining the number of dice). An increase of the difficulty level compared to the difficulty level the children were faced with themselves was evaluated as an intention to hurt that other child, and,

thus, as a behavioral measure of aggressive behavior (Baron & Richardson, 1994). The aggression score ranged from 0 (*no increase of the difficulty level*) to 5 (*highest increase of the difficulty level*). This behavioral measure of aggression was only used in Paper 1 to validate the behavioral observation measure of anger regulation. Since this measure was only used at T3, it was not appropriate to include in the longitudinal analyses that were in the focus of Paper 2 and Paper 3.

4.3.2 Instruments for the Risk Factors of Aggression

Anger Regulation. Paper 1 focused on the construct of anger regulation as an affective developmental factor of aggression. To assess children's use of anger regulation, a behavioral observation measure was used at T1 and T3. The assessment at T1 was developed and conducted by Rohlf and Krahé (2015). At T3, it was adapted to remain age appropriate. At both measurement points, the children were given a virtually impossible dexterity task that they were required to complete in a limited time to receive a present. Such procedures are supposed to induce anger (e.g., Helmsen & Petermann, 2010). Children's use of five different anger regulation strategies during the anger-eliciting task was coded based on recorded video material. The regulation strategies included four maladaptive (visual and verbal focus on the frustrating stimuli, venting the anger, and resignation) and one adaptive form of anger regulation (solution orientation).

Theory of Mind. Paper 2 focused on deficits in theory of mind as a cognitive risk factor of aggression. To assess children's theory of mind, a cartoon-based procedure was used at T1 and T2. The cartoons were taken from studies of Sebastian et al. (2012). The children were presented pictures depicting a story with an open end. Then, the children were shown two further pictures from which they had to select the one that present the correct ending (score 0 for an incorrect, score 1 for a correct response). There were 12 cartoons in total: six for affective, and six for cognitive theory of mind, respectively.

Physical Attractiveness. Paper 3 focused on children's physical attractiveness as an appearance-related factor of aggression. At T1 and T3, independent student raters were asked to evaluate the children's physical attractiveness based on facial photos. The response scale ranged from 1 (*not at all attractive*) to 7 (*very attractive*). For each child, an attractiveness score for T1 and T3 was obtained by aggregating the corresponding ratings. This procedure is commonly used in attractiveness research (e.g., Rosen et al., 2011).

Peer Problems. Children's problems with peer relationships (peer problems) were used as a validation construct in Paper 1 and as an underlying mechanism between physical attractiveness and aggressive behavior in Paper 3. At all three measurement points, three sources provided information on children's peer problems: the teachers, the children themselves, and their parents. For the teacher-report, three items were used with the three response options 1 (*not true*), 2 (*somewhat true*), and 3 (*certainly true*). Two items were taken from the Peer Problems subscale of the Strength and Difficulties Questionnaire (SDQ; Goodman, 1997), the third item was self-constructed to take the teacher's special role in children's school context into account. For the parent-report, all five items of the Peer Problems subscale of the SDQ were used with the same response options as the teacher-report. For the child-report, eight items at T1 and T2, and seven items at T3 were used. The items at T1 and T2 were taken from the Social Integration subscale of the Questionnaire on Social and Emotional Experiences at School of Elementary School Children (FEESS; Rauer & Schuck, 2003, 2004) and the Peer Acceptance subscale of the German version of the Harter-Scales (Asendorpf & van Aken, 1993). The response options were 1 (*no*) and 2 (*yes*). The items at T3 were taken from the FEESS and the response options were 1 (*no*), 2 (*rather no*), 3 (*rather yes*), and 4 (*yes*). For the validation procedure in Paper 1, all reports were used. However, in Paper 3, only the child-report was used as a measure of peer problems because the focus was on children's experienced problems with peers rather than an external assessment based on observation. An external assessment may be confounded with children's level of physical attractiveness (for a meta-analytic review, see Langlois et al., 2000).

4.4 Overview of Statistical Analyses

In all empirical papers, statistical analyses for testing the proposed hypotheses were done using structural equation modelling in Mplus (Muthén & Muthén, 1998-2015). However, the papers differ in their procedural approaches that were applied to examine the specific research questions. Table 4 provides an overview of the model specifications used in the empirical papers.

In Paper 1, latent structural equation modelling was used. Most indicators of the latent factors were of ordinal nature, and, therefore, diagonally weighted least squares (WLSMV) estimations were used instead of traditional maximum likelihood estimations (Li, 2016). Using WLSMV suggests the application of multiple imputation to deal with missing values

appropriately (Asparouhov & Muthén, 2010a, 2010c). Thus, in Paper 1, latent structural equation modeling was applied using the WLSMV estimator and multiple imputation.

Table 4. Overview of model specifications in the empirical papers

	Paper 1	**Paper 2**	**Paper 3**
SEM	all: latent	theory of mind: latent, aggression: manifest with latent random intercept	all: manifest, aggression with latent random intercept
Estimator	WLSMV	MLR	MLR
Handling of missing values	multiple imputation	FIML	FIML

Notes. SEM = structural equation model, WLSMV = diagonally weighted least squares, MLR = robust maximum likelihood, FIML = full information maximum likelihood.

In Paper 2, the predictor variable of theory of mind was also modelled latently, but aggressive behavior was included manifestly with the inclusion of a latent random intercept for the construct of aggressive behavior. Random intercepts partial out stable between-person differences and, thus, provide more precise examinations of within-person processes (Hamaker, Kuiper, & Grasman, 2015). In addition, all indicators were of metric nature, and, therefore, robust maximum likelihood (MLR) estimations were used (Muthén & Muthén, 1998-2015). Using MLR allows to use Full Information Maximum Likelihood (FIML) procedures to deal with missing values (Enders & Bandalos, 2001). Thus, in Paper 2, structural equation modeling was applied with latent modeling of the predictor and manifest modeling of the outcome, including a latent random intercept for the outcome, and using the MLR estimator and FIML.

In Paper 3, a procedure similar to that used in Paper 2 was applied. The central outcome of aggressive behavior was again used manifestly with the inclusion of a latent random intercept. However, since the central predictor of physical attractiveness was a one-item construct, latent modelling for this variable was not applicable. In consequence, this construct was included as a manifest variable. As all constructs were of metric nature, MLR estimators

were used once again, and, consequently, FIML was used to handle missing values. Thus, in Paper 3, manifest structural equation modeling was applied including a latent random intercept for the outcome, and using the MLR estimator and FIML.

5

Empirical Studies

5.1 Using Behavioral Observation for the Longitudinal Study of Anger Regulation in Middle Childhood (Paper 1)

Abstract

Assessing anger regulation via self-reports is fraught with problems, especially among children. Behavioral observation provides an ecologically valid alternative for measuring anger regulation. The present study uses data from two waves of a longitudinal study to present a behavioral observation approach for measuring anger regulation in middle childhood. At T1, 599 children from Germany (6-10 years old) were observed during an anger eliciting task, and the use of anger regulation strategies was coded. At T2, 3 years later, the observation was repeated with an age-appropriate version of the same task. Partial metric measurement invariance over time demonstrated the structural equivalence of the two versions. Maladaptive anger regulation between the two time points showed moderate stability. Validity was established by showing correlations with aggressive behavior, peer problems, and conduct problems (concurrent and predictive criterion validity). The study presents an ecologically valid and economic approach to assessing anger regulation strategies *in situ*.

Keywords: anger regulation, observation, childhood, validation, longitudinal study.

Kirsch, F., Busching, R., Rohlf, H., & Krahé, B. (2017). Using behavioral observation for the longitudinal study of anger regulation in middle childhood. *Applied Developmental Science*. Advance online publication. doi:10.1080/10888691.2017.1325325

In everyday life, humans are constantly faced with different emotional states of positive and negative valence (Trampe, Quoidbach, & Taquet, 2015). As one of the basic emotions (Ekman & Friesen, 1971), anger has received special attention, not least because it has been associated with aggressive behavior in many studies (e.g., Wittmann, Arce, & Santisteban, 2008). In childhood in particular, anger is a very common and intense emotion (von Salisch, 2000). Children report greater difficulties (Waters & Thompson, 2014) and lower self-efficacy (Zeman & Shipman, 1997) in dealing with anger compared to other negatively valenced emotional states, such as sadness. To avoid the social and behavioral problems associated with unfiltered anger expression, children need to learn to deal with their anger in an appropriate way. That is, they have to learn the adaptive regulation of the emotional state of anger. There is a broad range of problems associated with maladaptive anger regulation, including aggressive behavior (e.g., Helmsen & Petermann, 2010), peer problems (e.g., von Salisch, Zeman, Luepschen, & Kanevski, 2014), and conduct problems (e.g., Morris, Silk, Steinberg, Terranova, & Kithakye, 2010).

Given the critical role of maladaptive anger regulation as a potential risk factor for behavioral and peer problems, studying the development of anger regulation on the basis of longitudinal designs is an important task. In the present research, we propose and validate a behavioral observation method that lends itself to the longitudinal study of anger regulation in middle childhood by exposing children to an anger-eliciting task at successive points in time and observing their regulation strategies. This methodological approach has the advantage of yielding information about a child's anger regulation in a real-life situation. At the same time, it faces the challenge of developing conceptually equivalent, but age-adapted versions of the anger-eliciting task. This is the basis for comparing anger regulation strategies over time. In earlier work, we assessed children's anger regulation strategies through observation and related them to aggressive behavior and peer problems cross-sectionally (Rohlf & Krahé, 2015). Based on this work, we designed an age-adapted task that was used to study the development of anger regulation in relation to developmental problems at a second assessment 3 years later. Specifically, we considered the role of maladaptive anger regulation as a predictor of aggressive behavior, peer problems, and conduct problems.

Measuring Anger Regulation in Middle Childhood

Past research on anger regulation in middle childhood has mostly used self-reports to measure children's anger regulation skills (Aldao, Nolen-Hoeksema, & Schweizer, 2010), but this

procedure often involves difficulties (Underwood, 1997). Self-reported anger regulation may be biased as children may not be aware of their actual emotions or the way in which they try to regulate them. In actual anger eliciting situations, the emerging emotions may influence information processing (Berkowitz, 2012) and, consequently, the use of regulation strategies. Thus, self-reports may reflect children's knowledge of regulation strategies rather than their actual behavior. As a result, self-reports of the use of potential anger regulation strategies in hypothetical situations are unlikely to reflect the use of anger regulation strategies in real life. For example, in a study conducted by Parker et al. (2001), 6- to 11-year-old children generated fewer strategies in real-life anger eliciting situations than they mentioned for comparable hypothetical situations.

Given these limitations of self-reports, there is a need for more ecologically valid instruments that assess anger regulation in situations where anger is actually experienced. In order to address this task, Rohlf and Krahé (2015) developed and validated a behavioral observation method for children in the age group of 6 to 10 years. Anger was induced by having the children work on a virtually unsolvable dexterity task (i.e., a tower-building task). Anger regulation was assessed by means of a coding scheme that categorized children's specific regulation strategies, including visual and verbal focus on frustrating stimuli, venting the anger, resignation, and solution-oriented behavior. Strategies were classified as adaptive or maladaptive with respect to the prevention or promotion of negative interpersonal outcomes, such as aggressive behavior and social rejection, as explained in detail in Rohlf and Krahé (2015).

The *in-situ* method provides ecologically valid information about children's anger regulation strategies and avoids several of the problems of self-reports (Rohlf & Krahé, 2015; Kirsch, Rohlf, & Krahé, 2015). However, it poses challenges for the use in longitudinal studies, in particular with regard to the comparability of the anger-eliciting task. One way to address this problem is to use the exact same task (Wohlwill, 1973). However, presenting the same task repeatedly could lead to higher reactivity (including possible memory and training effects), which threatens the internal validity of the assessment (Schaie & Hofer, 2001). For instance, children who are exposed to exactly the same anger-arousing situation for the second time may use more effective strategies to deal with their emotions as they are familiar and experienced with this setting. Additionally, using the same measure at different ages may not be appropriate because children become more cognitively and motorically skilled (e.g., Bartolotta & Shulman, 2010). As they get older, children may understand that the tower-building task is in fact

impossible to complete, and the nature of the task may no longer be age-appropriate as they have outgrown playing with bricks. On the other hand, introducing a novel task can be potentially problematic as differences between the tasks may undermine the comparability of findings over time.

Thus, establishing the equivalence of anger-eliciting tasks used in the longitudinal study of anger regulation through behavioral observation is a methodological challenge. To maximize equivalence, the nature of the task should remain similar over time and require similar skills yet take into account maturational changes in behavior and interest.

The Present Study

The present study aimed to further develop the behavioral observation method of anger regulation presented by Rohlf and Krahé (2015) for use in longitudinal research studies. A major goal of our study was the development of a new observational measure to examine anger regulation longitudinally in middle childhood. The anger-eliciting task was selected to be appropriate for both boys and girls, that is without stereotypical gender preferences or connotations. This was deemed important to be able to evaluate potential gender differences irrespective of the characteristics of the specific anger-eliciting task. A few studies have revealed gender differences in the regulation of anger (e.g., Underwood, Coie, & Herbsman, 1992) and in the evaluation of the adaptiveness of specific anger regulation strategies (e.g., Waters & Thompson, 2014). In addition, we sought to demonstrate, in an exemplary fashion, the challenges and advantages of designing equivalent measures to study developmental processes over time.

We designed an anger-eliciting task that was conceptually equivalent to the tower-building task reported in Rohlf and Krahé (2015), but adapted to the children's improved cognitive and motor skills 3 years after the initial assessment. We provided support for the equivalence of both assessments and for the validity of our new, adapted task. In addition, we determined the stability of anger regulation over the course of 3 years. The structural equivalence of both tasks was established by assessing longitudinal measurement invariance between the observational measures at Time 1 (T1) and Time 2 (T2). For comparing regression slopes longitudinally, it is essential that the latent factors for both time points have the same unit of measurement (Chen, 2007). Thus, we aimed to establish at least metric invariance.

A high correlation between the measures at T1 and T2 indicates construct stability, that is, consistency of anger regulation strategies across time. As emotion regulation is assumed to be a stable individual difference variable in middle childhood (Cole, Michel, & Teti, 1994) we expected a substantial correlation between our two measures.

To assess validity, we examined the relations between the T1 and T2 assessments of maladaptive anger regulation and aggressive behavior, peer problems, and conduct problems (criterion validity). These constructs were correlated with anger regulation in past research. For example, aggressive children used more maladaptive anger regulation strategies than did nonaggressive children, including focusing on frustrating stimuli and venting the anger (Helmsen & Petermann, 2010). Another study found cross-sectional and longitudinal relations between anger dysregulation (venting) and externalizing behavior (Morris et al., 2010). Regarding the peer context, anger regulation was identified as an important predictor of having reciprocal friendships (von Salisch et al., 2014).

The following hypotheses were examined in our study: (1) The new anger-eliciting task developed at T2 is conceptually equivalent to the task used at T1, as reflected in metric invariance between the two tasks. (2) Anger regulation observed at T1 is substantially correlated with anger regulation observed at T2 based on an age-adapted version of the anger-eliciting task (construct stability). (3) Anger regulation observed at T1 shows prospective associations and anger regulation at T2 shows concurrent associations with aggressive behavior, peer problems, and conduct problems (criterion validity).

Based on conceptual considerations as well as past empirical research (e.g., Morris et al., 2010), we assumed the proposed relations to hold for both boys and girls, and we expected no age differences. To address these assumptions, we tested the invariance of our measures across gender and age groups, and examined gender and age group differences in the proposed associations between our measures of anger regulation and the validation constructs.

Method

Participants and Procedure

The sample consisted of 599 children from Germany (50.8% girls) who took part in the T1 assessment of a larger longitudinal study on intrapersonal developmental risk factors in childhood and adolescence (see Appendix I-A, for a more detailed description). They were

between 6 and 10 years old at T1 (M = 8.12 years, SD = 0.92). Of these, 554 children (50.2% girls) took part in the T2 measurement about three years later, which corresponds to a high retention rate of 92.5%.[1] The time interval between T1 and T2 was on average 2.77 years (SD = 0.19). Thus, the children were between 9 and 13 years old at T2 (M = 10.81 years, SD = 0.90). To avoid a reduction in sample size, all 599 T1 participants were included in the study, and missing data were handled using multiple imputation, as described in the following section.

The sample was recruited from 33 community elementary schools representing a variety of rural and urban areas, and socio-economic backgrounds. In total, 33.9% of the mothers and 36.1% of the fathers reported holding a university degree, 22.9% and 13.6% had university entrance qualification, 41.6% and 48.9% had a vocational-level qualification, and 1.6% and 1.4% had no or a low level of school qualification. Because the study was conducted in a part of Germany where the ethnic diversity is low (i.e., mostly Caucasian), we did not explicitly ask about ethnic background. However, we asked the parents which language they primarily spoke with their children. The vast majority of the parents (94%) reported speaking exclusively German with their children, and only 0.4% reported exclusively speaking a foreign language.

At both waves, data collection took place at the participants' school and was conducted by trained project staff. The materials and procedure were approved by the Ethics Committee of the authors' university as well as the Ministry of Education, Youth and Sport of the Federal State of Brandenburg, Germany, where the study was conducted. Active written consent was obtained from the parents, and the children provided assent before the start of the data collection.

Anger-Eliciting Tasks

To observe children's use of anger regulation strategies, two similar yet age-adapted anger-eliciting tasks were designed for T1 and T2. In both tasks, participants were given a virtually impossible dexterity task. At T1, the task involved building a tower with 10 wooden blocks on the basis of a photo placed in front of the child. The tower collapsed every time, since two of these blocks were slightly rounded on one side. The T1 task is described in detail in Rohlf and Krahé (2015). The task required the motor skills to balance the blocks to equilibrium. The new task developed for use 3 years later at T2 tapped into the same ability, but it was modified to be

[1] The data are part of a study that included three waves, in which the behavioral observation measure was employed at T1 and T3. Only data from these two waves are used in the present study. To avoid confusion, we refer to the two data waves as T1 and T2 for the purposes of the current analysis.

age-appropriate. The new task involved stacking seven dice to form a tower. The dice were of different sides, including two four-sided dice (tetrahedrons). Because of these two tetrahedrons, the dice tower task was practically impossible to accomplish (tetrahedrons only have one contact surface, which makes balancing the dice very difficult). A (photo-shopped) picture of a possible dice tower was placed in front of the children, but they were told that it just showed an example and they were free to stack the dice as they liked. The pictures used in the two tasks are presented in Appendix I-B. At both T1 and T2, there was a time restriction of 2 minutes and 40 seconds, determined in a pilot study as an appropriate time window for the display of anger regulation strategies. An hourglass (T1) or an electronic timer (T2) was placed on the table, so that the children could see the time running out. In addition, the children were told that they would get one of three presented gifts if they managed to complete the task in the given time. The desirability of the gifts was established by pilot tests. These gifts were also placed on the table. The experimenters sat behind the children, so the children would not be distracted by their presence and could concentrate on the task. All children were videotaped while working on the task.

Coding Anger Regulation Strategies

To assess the children's anger regulation strategies, their behavior during the anger-eliciting task was coded based on the categorization of adaptive and maladaptive anger regulation strategies. The development and classification of these strategies as adaptive or maladaptive is explained in detail in Rohlf and Krahé (2015). At both time points, four maladaptive anger regulation strategies (*visual focus on the frustrating stimuli*, *verbal focus on the frustrating stimuli, venting the anger*, and *resignation*) and one adaptive strategy (*solution orientation*) were coded for each child. Each strategy was represented by at least one specific codable behavior (subcategory) that was assumed to reflect the superordinate strategy (see Table I-1). For the majority of strategies, event-coding was used, counting the number of occurrences of each subcategory within the observation period (exceptions are described below). Coding was conducted by trained raters using the software Eudico Linguistic Annotator (ELAN; Wittenburg, Brugman, Russel, Klassmann, & Sloetjes, 2006).

The definition of the four maladaptive strategies was the same at both time points. *Visual focus on the frustrating stimuli* included eye movement toward potential frustrating objects, that is, the unreachable gifts and the running timer. *Verbal focus on the frustrating stimuli* included any verbal comments referring to potentially frustrating objects, that is, the running time, the

unreachable gifts, the task itself, and the self (e.g., "I can't do it"). *Venting the anger* described any anger expressive behavior, including verbal expressions (e.g., swearing), facial and gestural expressions (e.g., clenching one's fist), and rough handling of the material (e.g., smashing the blocks or dice on the table). *Resignation* included refusing to continue working on the task for at least 3 seconds.

Table I-1. Coding scheme and inter-rater reliability

Strategy	α_{T1}	α_{T2}	Sub-categories
Visual focus on frustrating stimuli	.71	.84	– Looking at the timer – Looking at the gifts
Verbal focus on frustrating stimuli	.92	.85	– Talking negatively about the time – Talking negatively about the gifts – Talking negatively about the task – Talking negatively about the self
Venting the anger	.73	.74	– Verbal expression of anger – Anger expression in facial expression and gesture – Handling the material roughly
Resignation	.99	.92	– Refusing to continue for at least 3 sec (0 vs. 1)
Solution orientation	.79		T1 – Testing an alternative strategy – Balancing out the tower (in sec) – Working in a focused way (1-4)
		.89	T2 – Testing an alternative strategy (0-5) – Balancing out the tower – Working in a focused way (1-4) – Rearranging the dice – Following the example picture (0 vs. 1)

Note. α = Krippendorff's alpha.

Due to some changes of the anger-eliciting task in material and procedure, the only adaptive strategy *solution orientation* was defined slightly differently at both time points: At

T1, the sub-category "Testing an alternative strategy" was event-coded, counting the numbers of new attempts to complete the task. At T2, we counted the first occurrences of five predefined strategies, resulting in a scale ranging from 0 (s*howed none of these strategies*) to 5 (*showed all of these strategies*). For the sub-category "Balancing out the tower", the time (in seconds) that the children spent trying to balance the tower was measured at T1. The dice tower at T2 was far less stable, so the tower collapsed much more often, and the duration of each balancing attempt could not be measured accurately. Instead, we counted the *number* of attempts to balance the dice tower (event-coded). At both time points, the sub-category "Working in a focused way" was rated by the coder on a scale from 1 (*very little engagement with the task*) to 4 (*very much engagement with the task*). Since at T2 the children did not have to copy the tower depicted on the picture, but were free to stack the dice as they wanted, they had more opportunities to show solution-oriented behavior. Thus, the number of times they rearranged the dice was event-coded as an additional category, and we also coded whether the children followed the example picture (0 vs. 1).

Sum scores of the corresponding sub-categories were calculated for the three maladaptive strategies v*isual focus on frustrating stimuli*, *verbal focus on frustrating stimuli*, and *venting the anger*. A dichotomized score was created for *resignation* (0 vs. 1)[2]. For the adaptive strategy *solution orientation*, the scaling of the subordinate categories differed, as explained above. Therefore, scores for each sub-category were first *z*-standardized and then averaged into an overall score.

A sub-sample of videos (n = 121 at T1 and n = 120 at T2) were double-coded by an independent rater to estimate the reliability of the coding process. The specific coding scheme as well as reliability information for both time points are presented in Table I-1. We calculated Krippendorff's alpha as a measure of reliability of the coding. This coefficient is appropriate for coded data and determines the agreement between coders (Hayes & Krippendorff, 2007). Krippendorff's alphas ranged from .71 to .99 at T1 and from .74 to .92 at T2, respectively, indicating acceptable to good interrater agreement for both time points.

[2] Originally, event-coding was used for resignation at both time points. At T2, the maximum number of resignations was one. To improve model fit, resignation T2 was declared as a categorical variable in all latent analyses. For the sake of consistency over time and to be able to test measurement invariance, we retrospectively changed the T1 scoring to this dichotomized score. The data of only three children at T1 were affected by this modification, with their score changing from 2 to 1.

Emotional Reactions to the Anger-Eliciting Task

To check the success of our task in eliciting angry feelings in contrast to other negative emotions, children rated their experience of anger and sadness during the tower-building task on two items using a scale from 1 (*not at all angry/sad*) to 3 (*very angry/sad*). At T1, this assessment took place immediately after the tower-building task. At T2, it was placed after the assignment of the number of dice to the other child as a measure of aggressive behavior (explained in the following section).

Validation Constructs

To establish criterion validity, we assessed several constructs at T1 and T2 that are assumed to be associated with maladaptive anger regulation. In particular, we considered children's aggressive behavior based on three independent sources of information (teacher-report, child-report, and *in-situ* behavior), peer problems (teacher-, child-, and parent-report), and conduct problems (parent-report). These constructs (with the exception of *in-situ* behavior) were operationalized using Likert-type scales with two to five response options, that is they were measured at an ordinal level of measurement. Thus, we calculated Ordinal alpha instead of traditional Cronbach's alpha as an estimation of reliability (Gadermann, Guhn, & Zumbo, 2012).

Aggressive behavior: Teacher-report. At both data waves, teachers indicated the frequency of children's aggressive behavior in the last 6 months. We used six items adapted from the Children´s Social Behavior Scale - Teacher Form (CSBS-T; Crick, 1996; e.g., "How often did this child hit, shove, or push peers"; $\alpha_{t1} = .94$, $\alpha_{t2} = .95$). The response scale ranged from 1 (*never*) to 5 (*daily*).

Aggressive behavior: Child-report. At T2, the children were presented the same six CSBS-items as the teachers in a reformulated, age-adapted version (e.g., "How often did you hit, shove, or push other children"; $\alpha = .74$). The response scale ranged from 1 (*never*) to 4 (*almost daily*).

Aggressive behavior in situ. At T2, a behavioral measure of aggression was developed that was derived from the dice-stacking task. Immediately after the task, the children were told that they could decide how many dice another child in another school would receive to complete the same task. Participants were told that they could choose any number between 2 and 12 dice

and were reminded that they had been given seven dice. The number of additional dice (>7) assigned to the alleged other child was taken as the measure of aggressive behavior. This operationalization of aggressive behavior is based on the same rationale as the Tangram Help/Hurt Task by Saleem, Anderson, and Barlett (2015). Because the task gets more difficult the more dice have to be used to build the tower, assigning a greater number of dice to another child than the participant had to handle him/herself can be viewed as reflecting an intention to harm, thus meeting the definition of aggressive behavior (Baron & Richardson, 1994). Accordingly, we transformed the raw number of dice into a scale using 7 dice as the reference point. All assignments of equal or less than 7 received a score of zero since assigning as many or fewer dice to the other child than they had received themselves represents a nonaggressive response. The resulting aggression scale ranged from 0 (*7 dice or fewer*) to 5 (*maximum number of 12 dice*).

Peer problems: Teacher-report. At T1 and T2, teachers rated each child on the degree of experienced peer problems, using three items (α_{t1} = .85, α_{t2} = .87). Two were taken from the Peer Problems subscale of the Strength and Difficulties Questionnaire (SDQ; Goodman, 1997; "is picked on or bullied by other children" and "is generally liked by other children", reverse coding). The third item was self-constructed for taking into account the school context ("is often excluded when classmates play together at break time"). The response scale ranged from 1 (*not true*) to 3 (*certainly true*).

Peer problems: Child-report. At both time points, children were asked to report on their problems with peers in their class. At T1, peer problems were assessed using eight items (α = .80). Of these, five items were taken from the Social Integration subscale of the Questionnaire on Social and Emotional Experiences at School of Elementary School Children (Fragebogen zur Erfassung emotionaler und sozialer Schulerfahrungen, FEESS; Rauer & Schuck, 2003; e.g., "The other children often laugh at me"), and three items were taken from the Peer Acceptance subscale of the German version of the Harter-Scales (Asendorpf & van Aken, 1993; e.g., "I am liked by other children", reverse coding). The response options at T1 were 1 (*no*) and 2 (*yes*). At T2, peer problems were assessed with seven items of the FEESS (including the same as at T1; α = .87). The response options at T2 were 1 (*no*), 2 (*rather no*), 3 (*rather yes*), and 4 (*yes*).

Peer problems and conduct problems: Parent-report. At both time points, parents completed the Peer Problems (α_{t1} = .79, α_{t2} = .83) and the Conduct Problems subscales (α_{t1} = .79, α_{t2} = .79; e.g., "often lies or cheats") of the SDQ (Goodman, 1997). The response scale ranged from 1 (*not true*) to 3 (*certainly true*).

Data Analysis Plan

We used Mplus (Version 7.3, Muthén & Muthén, 1998-2015) for our analyses. Due to the categorical nature of most indicators, the WLSMV estimator was used in all latent analyses (Li, 2015). Potential age and gender differences in latent analyses were examined by applying Wald tests. Age groups were defined by splitting the sample in a younger and an older sub-sample by median (*Md* = 8.02 at T1).

Only 7.5% of the T1 sample did not participate at T2, and few differences were found between the participants who remained in the sample and those who dropped out. Dropouts were older (t[597] = -2.18, p = .030, d = 0.34, 95% CI [0.03, 0.65]), and were described by their teachers as more aggressive at T1 (t[46.86] = -2.02, p = .049, d = 0.35, 95% CI [0.04, 0.66]). In terms of missing responses, a total of 13% of data were missing, ranging from 0% to 34.2% across the examined variables. The highest missing rate of 34.2% was for teacher reports at T2 since not all teachers returned the questionnaires.

To deal with the missing data and to reduce biases due to selective dropout, multiple imputation was used (Asendorpf, van de Schoot, Denissen, & Hutteman, 2014; Rubin, 1987). This was done under the missing at random (MAR) assumption. We created 60 multiply imputed datasets in 120 iterations using R 3.2.1 (R Core Team, 2015), the default settings of the mice 2.25 package (van Buuren & Groothuis-Ousdshoorn, 2011), and fully conditional specification (van Buuren, Brand, Groothuis-Oudshoorn, & Rubin, 2006). For the sake of consistency, we also imputed missing data at T1, resulting in slight deviations from analyses reported in previous studies using the T1 data (Kirsch et al., 2015; Rohlf, Busching, Krahé, in press; Rohlf & Krahé, 2015). In addition to the variables used in the present study, we also included auxiliary variables (the remaining SDQ subscales) in the imputation model to improve parameter estimation (Yoo, 2009). All analyses reported in the Results section were based on the imputed data set.

Results

Manipulation Check

The tower-building task was designed to induce angry feelings in the children in order to observe actual use of anger regulation strategies. At both time points, anger ratings were above the midpoint of the response scale, and the tasks elicited significantly more anger than sadness

(T1: M_{anger} = 2.31, SD_{anger} = 0.64 vs. $M_{sadness}$ = 1.84, $SD_{sadness}$ = 0.72; t[598] = 11.97, p < .001, d = 0.94, 95% CI [0.89, 1.00]; T2: M_{anger} = 2.22, SD_{anger} = 0.59 vs. $M_{sadness}$ = 1.65, $SD_{sadness}$ = 0.61; t[598] = 15.43, p < .001, d = 1.24, 95% CI [1.19, 1.29]). This indicates that the anger induction was successful. In addition, experienced anger was not correlated with participant's age (rs ≤ |-.02|, ps ≥ .643). This indicates that our task was able to elicit anger regardless of age. Regarding gender, boys reported more anger than did girls at T1 (M_{girls} = 2.25, SD_{girls} = 0.63 vs. M_{boys} = 2.38, SD_{boys} = 0.64; t[597] = 2.46, p = .014, d = 0.20, 95% CI [0.04, 0.36]), but there was no difference at T2 (t[597] = 0.97, p = .330, d = 0.08, 95% CI [-0.08, 0.24]).

As a manipulation check for the *in-situ* measure of aggressive behavior used at T2 (number of dice), a sub-sample of n = 76 randomly selected children were asked whether they thought the task was more difficult when more dice had to be used. The vast majority of these children (86%) answered in the affirmative. This result supports the assumption that participants were aware of the harmful nature of assigning more dice to the alleged other child, suggesting that assigning a higher number of dice than they had received themselves can be interpreted as a measure of aggressive behavior, defined by the intention to harm (Krahé, 2013).

Descriptive Statistics, Gender and Age Differences, Factor Analysis, and Correlations

Descriptive statistics of the use of anger regulation strategies at T1 and T2 as well as tests for gender differences are presented in Table I-2. *Visual focus* and *venting the anger* were the most frequently used strategies, *resignation* had the lowest frequency. Few gender differences emerged in the use of anger regulation strategies: Boys used more *visual focus* than did girls at T1 (t[597] = 3.19, p = .002, d = 0.26, 95% CI [0.10, 0.42]), and more *venting the anger* at T2 (t[597] = 2.57, p = .010, d = 0.21, 95% CI [0.05, 0.37]).

Table I-3 presents stability information, factor loadings, bivariate correlations of the anger regulation strategies at T1 and T2, and their correlations with age. Stabilities over the course of the 3 years between T1 and T2 were low to moderate, rs ranging from .12 to .29. *Solution orientation* showed the lowest stability due to the redefinition of that indicator, as explained above. Pearson's correlation coefficients among the strategies were low to moderate at both time points. Age at T1 was negatively correlated with *visual* and *verbal focus*, and positively with *solution orientation* at both time points, suggesting that maladaptive anger regulation decreases with age. The measurement model for maladaptive anger regulation at T2 showed a good fit after freeing residual covariances between the manifest indicators *venting the*

anger and *verbal focus*, *resignation* and *solution orientation*, and *resignation* and *visual focus* ($\chi^2[2] = 3.88$, $p = .144$; RMSEA = .03; WRMR = 0.25; CFI = .99; TLI = .97). The measurement model for maladaptive anger regulation T1 also showed a good fit after freeing residual covariances between *visual focus* and *solution orientation*, and *resignation* and *solution orientation* ($\chi^2[3] = 11.84$, $p = .008$; RMSEA = .07; WRMR = 0.59; CFI = .99; TLI = .96). The factor loadings indicate that the maladaptive regulation strategies were positively associated with each other, whereas the single adaptive strategy *solution orientation* was negatively associated with the other strategies. This confirms the theoretical classification of the strategies as adaptive and maladaptive.

Descriptive statistics, gender differences, stabilities and correlations of the validation constructs with age are presented in Appendix I-C. Stability of the constructs were low to moderate across the time interval of 3 years (*r*s ranging from .29 to .64, *p*s < .001). All constructs were positively correlated with each other. The significant cross-sectional correlations of our *in-situ* measure of aggressive behavior (number of dice) with child- ($r = .22$, $p < .001$, 95% CI [.14, .30]) and teacher-reported ($r = .15$, $p < .001$, 95% CI [.07, .23]) aggressive behavior provide further support for its validity as a measure of aggressive behavior.

Measurement Invariance

To demonstrate the equivalence of the two tasks measuring anger regulation at T1 and T2, longitudinal measurement invariance was tested (Millsap & Cham, 2012). Table I-4 presents a summary of the corresponding analysis. A change of < -.010 in CFI, and < .015 in RMSEA was interpreted as supporting the invariance assumption (Chen, 2007). For establishing configural and higher levels of invariance, the residual covariance of the strategy *visual focus* at T1 and T2 was freed to improve model fit. Full invariance across all indicators could only be established at the configural level, but not for the higher levels. Steenkamp and Baumgartner (1998) recommend testing partial measurement invariance when full invariance cannot be applied. Thus, for further invariance tests, the strategy *solution orientation* was freed between the two time points, since this strategy had changed in definition over time. In the end, the highest level of invariance for our proposed behavioral observation method was partial metric invariance, as indicated by changes in fit indices below the critical threshold (ΔCFI = -.004; ΔRMSEA = -.001). That is, the factor loadings of all indicators – except for the strategy *solution orientation* – could be constrained to be equal across a time interval of 3 years. Additionally

Table I-2. Descriptive statistics and gender differences of anger regulation strategies

	T1					T2			
	Range	Total *M* (*SD*)	Girls *M* (*SD*)	Boys *M* (*SD*)	Gender difference	Range	Total *M* (*SD*)	Girls *M* (*SD*)	Boys *M* (*SD*)	Gender difference
Visual focus	0 - 39	4.07	3.60	4.55	$t = 3.19$**	0 - 14	3.45	3.38	3.52	$t = 0.58$
		(3.67)	(3.12)	(4.11)	$d = 0.26$		(2.98)	(2.93)	(3.02)	$d = 0.05$
Verbal focus	0 - 27	2.75	2.62	2.88	$t = 0.90$	0 - 10	0.81	0.74	0.90	$t = 1.20$
		(3.54)	(3.57)	(3.50)	$d = 0.07$		(1.63)	(1.56)	(1.69)	$d = 0.10$
Venting	0 - 22	4.33	4.36	4.29	$t = -0.22$	0 - 14	2.29	2.01	2.58	$t = 2.57$*
		(3.87)	(3.90)	(3.84)	$d = 0.02$		(2.73)	(2.54)	(2.88)	$d = 0.21$
Resignation	0 - 2	0.02	0.02	0.02	$t = 0.25$	0 - 1	0.07	0.07	0.07	$t = 0.29$
		(0.14)	(0.15)	(0.14)	$d = 0.02$		(0.26)	(0.25)	(0.26)	$d = 0.02$
Solution	-	0.00	-0.12	0.13	$t = 1.92$	-	-0.04	0.00	-0.08	$t = -1.94$
orientation[a]		(1.60)	(1.51)	(1.68)	$d = 0.16$		(0.51)	(0.50)	(0.51)	$d = 0.16$

Note. [a] Scores for solution orientation were *z*-transformed; $n_{girls} = 304$, $n_{boys} = 295$; *df* for all independent *t*-tests was 597.
* $p < .05$. ** $p < .01$.

Table I-3. Stabilities, correlations with age, factor loadings, and intercorrelations of anger regulation strategies

	Visual focus T1	Verbal focus T1	Venting T1	Resignation T1	Solution T1	Visual focus T2	Verbal focus T2	Venting T2	Resignation T2	Solution T2
Age T1	-.19***	-.21***	-.13**	-.06	.34***	-.14***	-.13**	-.02	.02	.10*
Visual focus T1		.34***	.11**	.08*	-.35***	**.25****	.11**	.04	.09*	-.09*
Verbal focus T1			.43***	.16***	-.43***	.07	**.29****	.23***	.02	-.05
Venting T1				.14***	-.27***	.05	.21***	**.22****	-.01	-.02
Resignation T1					-.32***	.05	.15***	.12**	**.17****	-.14***
Solution T1						-.09*	-.20***	-.13**	-.05	**.12****
Visual focus T2							.19***	.12**	.30***	-.21***
Verbal focus T2								.44***	.16***	-.16***
Venting T2									.15***	-.13**
Resignation T2										-.40***
Solution T2										
Factor loading	.36***	.91***	.47***	.38***	-.49***	.47***	.37***	.30***	.62***	-.44***

Notes. Solution = solution orientation; stabilities of strategies are highlighted in bold.

* $p < .05$. ** $p < .01$. *** $p < .001$.

constraining the intercepts of the indicators, a condition for scalar invariance, worsened the model fit, therefore partial scalar invariance could not be assumed.

In addition to measurement invariance over time, we tested measurement invariance across gender groups to examine the equivalence of the method for boys and girls (Widaman & Reise, 1997; see Table I-5). At both T1 and T2 full scalar invariance could be established (ΔCFI_{T1} = -.004, $\Delta RMSEA_{T1}$ = -.007; ΔCFI_{T2} = -.007, $\Delta RMSEA_{T2}$ = .001). That is, the factor loadings and the intercepts of all indicators could be constrained to be equal across boys and girls. Regarding the role of age, we tested the measurement invariance across the younger and older sub-sample of our study (see Table I-5). At T1, partial metric invariance could be established (the indicator verbal focus was freed; ΔCFI = .000, ΔRMSEA = -.014). At T2, full scalar invariance could be established (ΔCFI = -.009, ΔRMSEA = .012).

Table I-4. Measurement invariance of maladaptive anger regulation over time

Level of invariance	χ^2	*df*	TLI	WRMR	CFI	ΔCFI	RMSEA	ΔRMSEA
Configural	81.75	27	.928	0.999	.957	-	.058	-
Full metric	139.62	31	.876	1.447	.915	-.042	.076	.018
Full scalar	213.76	35	.820	1.802	.860	-.055	.092	.016
Full strict	381.18	39	.690	2.486	.732	-.128	.121	.029
Partial metric	89.55	30	.930	1.111	.953	-.004	.057	-.001
Partial scalar	137.34	33	.888	1.407	.918	-.035	.073	.016
Partial strict	190.41	36	.848	1.707	.879	-.039	.085	.012

Note. All χ^2-statistics are significant at $p < .001$.

Stability

To assess construct stability, a latent-state model was computed to test the correlation between the two latent factors of maladaptive anger regulation at T1 and T2. This model was run under the assumption of partial metric longitudinal invariance as explained above (for model fit information, see Table I-4). The latent correlation was r = .52 ($p < .001$, 95% CI [.37, .67]).

Table I-5. Measurement invariance of maladaptive anger regulation across gender and age

Level of invariance	χ^2	*df*	*p*	TLI	WRMR	CFI	ΔCFI	RMSEA	ΔRMSEA
Gender T1									
Configural	18.17	6	.006	.949	0.744	.985	-	.081	-
Full Metric	29.63	11	.002	.957	1.275	.977	-.008	.074	-.007
Full Scalar	37.69	16	.002	.966	1.459	.973	-.004	.067	-.007
Full Strict	84.11	20	.000	.919	2.320	.919	-.054	.103	.036
Gender T2									
Configural	4.94	4	.294	.986	0.302	.995	-	.027	-
Full Metric	13.62	9	.137	.971	0.804	.986	-.009	.036	.009
Full Scalar	21.30	14	.094	.970	1.035	.979	-.007	.037	.001
Full Strict	30.22	18	.035	.960	1.270	.964	-.015	.043	.006
Age T1									
Configural	12.462	6	.052	.976	0.599	.990	-	.059	-
Full Metric	44.60	11	.000	.906	1.582	.948	-.042	.101	.042
Full Scalar	121.03	16	.000	.798	2.704	.838	-.110	.148	.047
Full Strict	259.92	20	.000	.630	4.162	.630	-.208	.200	.052
Partial Metric	16.63	10	.092	.981	0.844	.990	.000	.045	-.014
Partial Scalar	95.65	14	.000	.820	2.224	.874	-.116	.139	.094
Partial Strict	212.31	17	.000	.645	3.479	.698	-.176	.196	.057
Age T2									
Configural	4.70	4	.320	.990	0.283	.996	-	.024	-
Full Metric	10.00	9	.351	.995	0.648	.994	-.002	.018	-.006
Full Scalar	19.37	14	.151	.980	0.975	.985	-.009	.030	.012
Full Strict	46.33	18	.000	.916	1.628	.924	-.061	.071	.041

Note. Partial invariance for age T1 established by freeing the indicator verbal focus.

This result indicates moderate stability of the latent factor of maladaptive anger regulation over the course of 3 years. The stability did not differ between girls and boys ($W = 0.03$, $p = .873$), but between the age groups ($W = 7.70$, $p = .006$). Younger children showed higher construct stability than did older children ($r_{younger} = .59$ vs. $r_{older} = .40$; all ps $< .001$).

Criterion Validity

Concurrent criterion validity of T1 and T2 maladaptive anger regulation was assessed by cross-sectional correlations with the criterion measures at both time points. Predictive criterion validity for behavioral observation at T1 was assessed by correlating it with the T2 criterion measures. All correlations are partial correlations, controlled for age and gender. All criterion constructs (with the exception of the *in-situ* measure of aggressive behavior, which was a single-item measure) were modelled as latent variables. The resulting structural equation models testing concurrent and predictive validity showed good to adequate model fit, with .912 < CFI < .997 and .035 < RMSEA < .073 (van de Schoot, Lugtig, & Hox, 2012). Table I-6 presents the results for concurrent and predictive validity. As expected, maladaptive anger regulation showed positive correlations with aggressive behavior, peer problems, and conduct problems, both cross-sectionally at both time points and prospectively from T1 to T2. The only exception was the nonsignificant prospective association of T1 anger regulation with T2 teacher-rated aggressive behavior, which needs to be interpreted in view of the fact that children were rated by different teachers at the two data waves. The strongest associations, with correlations mostly around $r = .40$, were found for the concurrent associations of anger regulation with the validation constructs at T2. Wald tests revealed no gender differences in these correlations (Ws ≤ 2.10, ps $\geq .147$), with the exception of the predictive validity of *in-situ* aggressive behavior ($W = 4.57$, $p = .033$). The prospective correlation between maladaptive anger regulation T1 and *in-situ* aggression T2 was only significant in girls ($r = .23$, $p < .001$), but not in boys ($r = -.01$, $p = .871$). No significant age group differences were found (Ws ≤ 3.31, ps $\geq .069$). In sum, the findings provide conclusive support for the validity of the behavioral observation measures.

Discussion

The aim of this study was to further develop the behavioral observation assessment of anger regulation proposed by Rohlf and Krahé (2015) for the use in longitudinal research on the

development of anger regulation in childhood and its role as a risk factor for aggressive behavior, peer problems, and conduct problems. To elicit anger, children in the age group of 6 to 10 years were given a practically unsolvable dexterity task at T1 (building a tower using wooden blocks). About 3 years later at T2, the same children were asked to work on a slightly different, but age-appropriate variation of the task that was supposed to be conceptually equivalent (building a tower by stacking dice). Both tower-building tasks were successful in eliciting angry feelings, which made them suitable for the observational study of anger regulation *in situ*.

Table I-6. Concurrent and predictive validity of the behavioral observation measures

	Maladaptive anger regulation					
	Concurrent T1		**Concurrent T2**		**Predictive T1 to T2**	
	r	95% CI	*r*	95% CI	*r*	95% CI
Aggressive behavior						
Teacher-report	.20***	[.08, .31]	.22*	[.02, .42]	.06	[-.06, .17]
Child-report	-[a]	-	.40***	[.17, .63]	.23***	[.10, .36]
In-situ behavior	-[a]	-	.22**	[.07, .36]	.11*	[.01, .20]
Peer problems						
Teacher-report	.17*	[.04, .30]	.42***	[.23, .62]	.27***	[.15, .40]
Child-report	.27***	[.15, .38]	.40***	[.23, .58]	.15*	[.03, .26]
Parent-report	.19**	[.06, .31]	.32**	[.12, .52]	.21**	[.08, .33]
Conduct problems						
Parent-report	.21***	[.09, .33]	.40***	[.19, .61]	.27***	[.14, .40]

Notes. [a] Not assessed at T1; all correlations are controlled for age and gender.
* $p < .05$. ** $p < .01$. *** $p < .001$.

The same coding scheme was used at both data waves to categorize specific anger regulation strategies. The reliability of our new task was good, as indicated by high interrater agreement. Only few gender differences in the use of regulation strategies were found, and these were in line with previous studies (e.g., boys vented their anger more frequently than did girls at T2; for a review, see Kerr & Schneider, 2008). Tests of measurement invariance revealed partial metric invariance as the highest level of longitudinal invariance. Metric invariance

indicates that the underlying latent factor has the same unit of measurement at both time points, which is a prerequisite for the comparison of regression slopes in longitudinal research (Chen, 2007). Furthermore, tests of measurement invariance across gender and age groups indicated that the factorial structure of both observation measures were similar for girls and boys, and younger and older children, respectively.

Regarding stability, the latent construct of anger regulation showed a significant and substantial correlation between the two time points. This result indicates moderate construct stability over time, and provides further evidence of the equivalence of the two measures. Moreover, the substantial stability is consistent with the conceptualization of emotion regulation as a stable individual difference variable (Cole et al., 1994).

To establish the validity of our behavioral observation measure, we assessed problematic interpersonal outcomes associated with maladaptive emotion regulation in general (criterion validity). Specifically, we examined the correlations of maladaptive anger regulation cross-sectionally (concurrent validity) as well as longitudinally (predictive validity) with aggressive behavior, peer problems, and conduct problems. These validation constructs were assessed by a multi-method approach drawing on self-, teacher-, and parent-reports as well as *in-situ* behavior. For both concurrent and predictive analyses, we were able to demonstrate positive and for the most part significant correlations between maladaptive anger regulation and the validation constructs. In addition, the correlations were similar in girls and boys, as well as in younger and older children in our sample. These findings support the validity of our measure as they are in line with a large body of research also showing that deficits in anger regulation are related to behavioral and peer problems (e.g., Dearing et al., 2002).

Overall, our study was successful in developing an equivalent of the original behavioral observation assessment of Rohlf and Krahé (2015) that takes developmental changes in the children's cognitive and motor skills as well as interests into account. Such equivalent measurement tools are required in the longitudinal study of anger regulation in childhood, covering substantial lengths of time.

Strengths and Limitations

We believe that our study has several strengths. It is based on the behavioral observation of anger regulation in a large sample of children who were assessed twice over a 3-year period. Our two assessments were found to be conceptually equivalent and both measures worked

equally well for boys and girls. A range of outcome measures was included to establish criterion validity of the measures at T1 and T2, using data from multiple informants. Moreover, the anger-eliciting task used at T2 had the advantage of yielding a behavioral measure of aggression *in situ*, the number of dice assigned to another child. This feature facilitates a direct mapping of anger regulation strategies onto aggressive responses within the same situational context.

At the same time, some limitations have to be acknowledged. First, since we used behavioral observation, our conclusions and inferences can be applied only to anger regulation strategies that can be inferred from overt behavior. Like any behavioral observation, our measure cannot capture mental strategies, such as cognitive reappraisal, that are not necessarily reflected in overt behavior. As children get older, cognitive strategies may become more and more important in emotion regulation processes (e.g., McRae et al., 2012). Therefore, for future studies we suggest a combination of behavioral observation with self-report measures to obtain a clearer picture of children's regulation strategies.

Second, the children were exposed to an arranged anger-eliciting situation that constrained the possible regulation strategies they could show. Specifically, children did not have the opportunity to select or modify the situation, two important regulation approaches (Gross, 1998, 2014). Another effective and frequently used adaptive strategy in natural situations is distraction from the frustrating stimuli (e.g., Denson, Moulds, & Grisham, 2012). However, in our tasks, the children had very limited opportunities to withdraw from the task and distract themselves.

Another limitation is that we only assessed criterion validity (concurrent and predictive). However, for the T1 task, construct validity could be demonstrated by relating it to the conceptually related construct of anger reactivity (Rohlf & Krahé, 2015). Given the construct validity of our T1 measure and the conceptualization of emotion regulation as a stable individual difference variable (Cole et al., 1994), the substantial temporal stability from T1 to T2 may be interpreted as an indication that our T2 task also has construct validity as a measure of anger regulation (Cronbach & Meehl, 1955). As a final limitation, we only assessed the experience of two emotional states, namely anger and sadness, during the tower-building task. Other affective states such as anxiety or joy should be considered in future studies.

Despite these limitations, the behavioral observation measure offers an age-appropriate, economic, and valid approach for assessing the use of anger regulation strategies in an actual anger-inducing situation in an equivalent way at different points of development in childhood. With carefully trained experimenters, the anger-eliciting tasks are easy to administer and can

be completed in a short time. The categories are well-defined and grounded in theory, capture a range of different strategies, and yield reliable codings, as indicated by the high inter-coder reliability. They add a methodological tool to the study of anger regulation that lends itself to the longitudinal analysis of both adaptive and maladaptive forms of anger regulation in middle childhood. At a more general level, they demonstrate an approach for designing conceptually and empirically equivalent measures to study developmental processes over time.

References

Aldao, A., Nolen-Hoeksema, S., & Schweizer, S. (2010). Emotion-regulation strategies across psychopathology: A meta-analytic review. *Clinical Psychology Review, 30*, 217–237. doi:10.1016/j.cpr.2009.11.004

Asendorpf, J. B., & van Aken, M. A. G. (1993). Deutsche Versionen der Selbstkonzeptskalen von Harter [German version of the Harter's self-concept scales]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 25*, 64–86.

Asendorpf, J. B., van de Schoot, R., Denissen, J. J. A., & Hutteman, R. (2014). Reducing bias due to systematic attrition in longitudinal studies: The benefits of multiple imputation. *International Journal of Behavioral Development*, *38,* 453–460. doi:10.1177/0165025414542713

Baron, R. A., & Richardson, D. S. (1994). *Human aggression* (2nd ed.). New York, NY, US: Plenum Press.

Bartolotta, T. E., & Shulman, B. B. (2010). Child development. In B. B. Shulman & N. C. Singleton (Eds.), *Language development. Foundations, processes, and clinical applications* (pp. 35–53). Sudbury, MA, US: Jones and Bartlett Publishers.

Berkowitz, L. (2012). A different view of anger: The cognitive-neoassociation conception of the relation of anger to aggression. *Aggressive Behavior, 38*, 322–333. doi:10.1002/ab.21432

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*, 464–504. doi:10.1080/10705510701301834

Cole, P. M., Michel, M. K., & Teti, L. O. D. (1994). The development of emotion regulation and dysregulation: A clinical perspective. *Monographs of the Society for Research in Child Development*, 59, 73-100.doi:10.2307/1166139

Crick, N. R. (1996). The role of overt aggression, relational aggression, and prosocial behavior in the prediction of children's future social adjustment. *Child Development, 67*, 2317–2327. doi:10.2307/1131625

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281–302. doi:10.1037/h0040957

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology, 17*, 124–129. doi:10.1037/h0030377

Dearing, K. F., Hubbard, J. A., Ramsden, S. R., Parker, E. H., Relyea, N., Smithmyer, C. M., & Flanagan, K. D. (2002). Children's self-reports about anger regulation: Direct and indirect links to social preference and aggression. *Merrill-Palmer Quarterly, 48*, 308–336. doi:10.1353/mpq.2002.0011

Denson, T. F., Moulds, M. L., & Grisham, J. R. (2012). The effects of analytical rumination, reappraisal, and distraction on anger experience. *Behavior Therapy, 43*, 355–364. doi:10.1016/j.beth.2011.08.001

Gadermann, A. M., Guhn, M., & Zumbo, B. D. (2012). Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. *Practical Assessment, Research & Evaluation, 17*(3), 1–13.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry, 38*, 581–586. doi:10.1111/j.1469-7610.1997.tb01545.x

Gross, J. J. (1998). The emerging field of emotion regulation: An integrative review. *Review of General Psychology, 2*, 271–299. doi:10.1037/1089-2680.2.3.271

Gross, J. J. (2014). Emotion regulation: Conceptual and empirical foundations. In J. J. Gross (Ed.), *Handbook of emotion regulation* (2nd ed., pp. 3–20). New York, NY, US: Guilford Press.

Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*, *1*, 77–89. doi:10.1080/19312450709336664

Helmsen, J., & Petermann, F. (2010). Emotionsregulationsstrategien und aggressives Verhalten im Kindergartenalter [Emotion regulation strategies and aggressive behavior of preschool children]. *Praxis der Kinderpsychologie und Kinderpsychiatrie, 59*, 775–791.

Kerr, M. A., & Schneider, B. H. (2008). Anger expression in children and adolescents: A review of the empirical literature. *Clinical Psychology Review*, *28*, 559–577. doi:10.1016/j.cpr.2007.08.001

Kirsch, F., Rohlf, H., & Krahé, B. (2015). Measuring anger regulation in middle childhood through behavioural observation: A longitudinal validation. *European Journal of Developmental Psychology*, *12*, 718–727. doi:10.1080/17405629.2015.1101375

Krahé, B. (2013). *The social psychology of aggression* (2nd ed.). Hove, UK: Psychology Press.

Li, C.-H. (2015). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods.* doi:10.3758/s13428-015-0619-7

McRae, K., Gross, J. J., Weber, J., Robertson, E. R., Sokol-Hessner, P., Ray, R. D., . . . Ochsner, K. N. (2012). The development of emotion regulation: An fMRI study of cognitive reappraisal in children, adolescents and young adults. *Social Cognitive and Affective Neuroscience, 7*, 11–22. doi:10.1093/scan/nsr093

Millsap, R. E., & Cham, H. (2012). Investigating factorial invariance in longitudinal data. In B. P. Laursen, T. D. Little, & N. A. Card (Eds.), *Handbook of developmental research methods* (pp. 109–127). New York, NY, US: Guilford Press.

Morris, A. S., Silk, J. S., Steinberg, L., Terranova, A. M., & Kithakye, M. (2010). Concurrent and longitudinal links between children's externalizing behavior in school and observed anger regulation in the mother–child dyad. *Journal of Psychopathology and Behavioral Assessment, 32*, 48–56. doi:10.1007/s10862-009-9166-9

Muthén, L. K., & Muthén, B. O. (1998-2015). *Mplus user's guide: Statistical analysis with latent variables* (7th ed.). Los Angeles, CA, US: Muthén & Muthén.

Parker, E. H., Hubbard, J. A., Ramsden, S. R., Relyea, N., Dearing, K. F., Smithmyer, C. M., & Schimmel, K. D. (2001). Children's use and knowledge of display rules for anger following hypothetical vignettes versus following live peer interaction. *Social Development, 10*, 528–557. doi:10.1111/1467-9507.00179

Rauer, W., & Schuck, K.-D. (2003). *FEESS 3-4: Fragebogen zur Erfassung emotionaler und sozialer Schulerfahrungen von Grundschulkindern dritter und vierter Klassen* [FEESS 3-4: Questionnaire for the assessment of social and emotional school experiences of elementary school children in third and fourth grade]. Göttingen, GER: Beltz Test GmbH.

R Core Team. (2015). R: A language and environment for statistical computing. Vienna, AUT: Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Rohlf, H., Busching, R., & Krahé, B. (in press). Longitudinal links between maladaptive anger regulation, peer problems, and aggression in middle childhood. *Merrill-Palmer Quarterly*.

Rohlf, H., & Krahé, B. (2015). Assessing anger regulation in middle childhood: Development and validation of a behavioral observation measure. *Frontiers in Psychology*, *6*, Article 453. doi:10.3389/fpsyg.2015.00453

Rubin, D. B. (1987). *Multiple Imputation for nonresponse in surveys*. New York, NY, US: John Wiley & Sons.

Saleem, M., Anderson, C. A., & Barlett, C. P. (2015). Assessing helping and hurting behaviors through the Tangram Help/Hurt Task. *Personality and Social Psychology Bulletin, 41*, 1345–1362. doi:10.1177/0146167215594348

Schaie, K. W., & Hofer, S. M. (2001). Longitudinal studies in aging research. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (5th ed., pp. 53–77). San Diego, CA, US: Academic Press.

Steenkamp, J.-B. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research, 25*, 78–107. doi:10.108 6/209528

Trampe, D., Quoidbach, J., & Taquet, M. (2015). Emotions in everyday life. *PloS one, 10*, e0145450. doi:10.1371/journal.pone.0145450

Underwood, M. K. (1997). Top ten pressing questions about the development of emotion regulation. *Motivation and Emotion, 21*, 127–146. doi:10.1023/A:1024482516226

Underwood, M. K., Coie, J. D., & Herbsman, C. R. (1992). Display rules for anger and aggression in school-age children. *Child Development, 63*, 366–380. doi: 10.2307/1131485

van Buuren, S., Brand, J. P., Groothuis-Oudshoorn, C. G., & Rubin, D. B. (2006). Fully conditional specification in multivariate imputation. *Journal of Statistical Computation and Simulation, 76*, 1049–1064. doi:10.1080/10629360600810434

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software, 45*, 1–67. doi:10.18637/jss.v045.i03

van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology, 9*, 486–492. doi:10.1080/17405629.2012.686740

von Salisch, M. (2000). *Wenn Kinder sich ärgern: Emotionsregulation in der Entwicklung*. [When children feel angry: Emotion regulation in development]. Göttingen, D: Hogrefe.

von Salisch, M., Zeman, J., Luepschen, N., & Kanevski, R. (2014). Prospective relations between adolescents' social-emotional competencies and their friendships. *Social Development, 23*, 684–701. doi:10.1111/sode.12064

Waters, S. F., & Thompson, R. A. (2014). Children's perceptions of the effectiveness of strategies for regulating anger and sadness. *International Journal of Behavioral Development, 38*, 174–181. doi:10.1177/0165025413515410

Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention. Methodological advances from alcohol and substance abuse research* (pp. 281–324). Washington, DC, US: American Psychological Association.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). *ELAN: A Professional Framework for Multimodality Research.* Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation, Genoa.

Wittmann, M., Arce, E., & Santisteban, C. (2008). How impulsiveness, trait anger, and extracurricular activities might affect aggression in school children. *Personality and Individual Differences, 45*, 618–623. doi:10.1016/j.paid.2008.07.001

Wohlwill, J. F. (1973). *The study of behavioral development*. New York, NY, US: Academic Press.

Yoo, J. E. (2009). The effect of auxiliary variables and multiple imputation on parameter estimation in confirmatory factor analysis. *Educational and Psychological Measurement, 69*, 929–947. doi:10.1177/0013164409332225

Zeman, J., & Shipman, K. (1997). Social-contextual influences on expectancies for managing anger and sadness: The transition from middle childhood to adolescence. *Developmental Psychology, 33*, 917–924. doi:10.1037/0012-1649.33.6.917

Appendix I-A

The participants were part of a larger sample of 1,657 children from 33 public elementary schools who participated in a longitudinal study on intrapersonal developmental risk factors in childhood and adolescence. Children were videotaped if their parents gave permission (n = 1,183). It was not possible to code all videos due to limited resources. Therefore, a sub-sample of 599 children was randomly chosen for the T1 coding procedure. The resulting sub-sample did not differ significantly on any of the T1 variables included in the present study (see Table I-7).

Table I-7. Differences between included and not included children on T1 variables

		Included		Not included		
	Range	*N*	*M* (*SD*)	*N*	*M* (*SD*)	Difference
Aggressive behavior						
Teacher-report	1 - 5	585	1.51 (0.68)	806	1.48 (0.66)	$t(1389) = -0.84$ $d = 0.04$
Peer problems						
Teacher-report	1 - 3	536	1.22 (0.34)	747	1.25 (0.37)	$t(1281) = 1.10$ $d = 0.08$
Child-report	1 - 2	598	1.18 (0.19)	1044	1.18 (0.21)	$t(1640) = 0.22$ $d = 0.01$
Parent-report	1 - 3	554	1.22 (0.30)	765	1.22 (0.31)	$t(1317) = 0.37$ $d = 0.02$
Conduct problems						
Parent-report	1 - 3	557	1.31 (0.32)	762	1.29 (0.30)	$t(1317) = -1.13$ $d = 0.06$

Note. All differences between the two groups are nonsignificant.

Appendix I-B

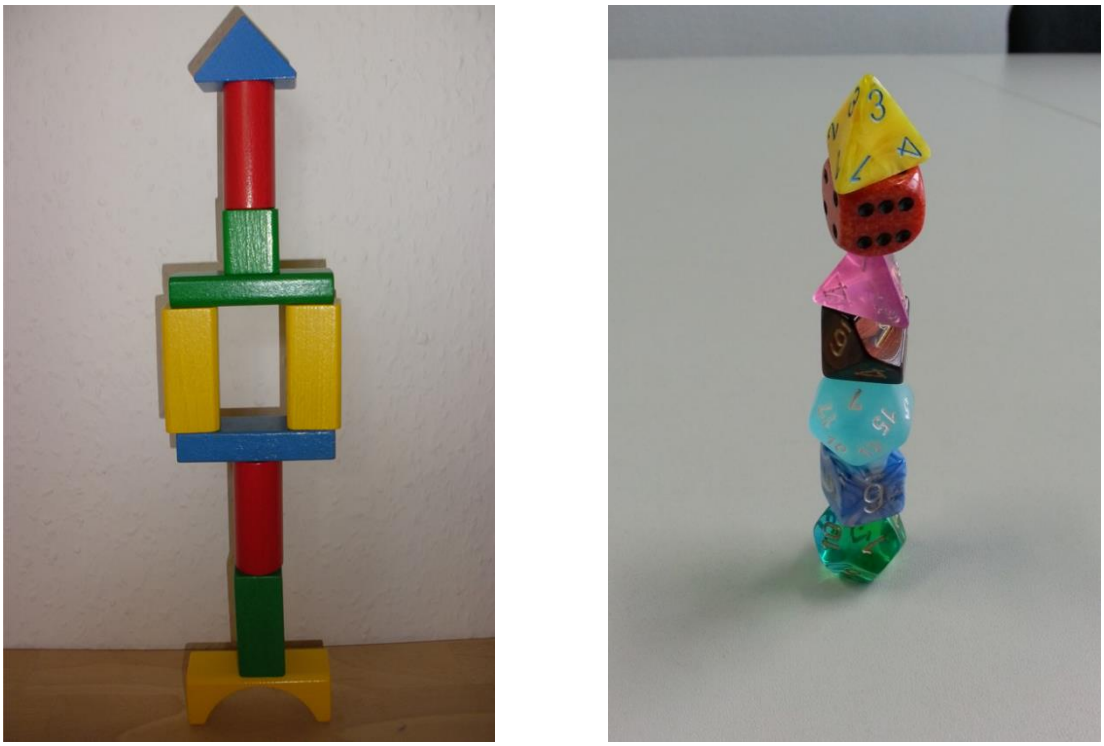

Figure I-1. Photos for the tower-building task used at T1 (left) and T2 (right)

Appendix I-C

Table I-8. Descriptive statistics and gender differences of validation constructs

		T1				T2			
	Range	Total *M* (*SD*)	Girls *M* (*SD*)	Boys *M* (*SD*)	Gender difference	Total *M* (*SD*)	Girls *M* (*SD*)	Boys *M* (*SD*)	Gender difference
Aggressive behavior									
Teacher-report	1-5	1.52 (0.68)	1.38 (0.58)	1.67 (0.75)	t = 5.30*** d = 0.43	1.57 (0.78)	*1.49* (0.70)	*1.65* (0.84)	t = 2.54* d = 0.21
Child-report[a]	1-4	-	-	-	-	1.29 (0.35)	*1.22* (0.29)	*1.36* (0.39)	t = 5.00*** d = 0.41
In-situ behavior[a]	0-5	-	-	-	-	0.73 (1.39)	0.74 (1.38)	0.72 (1.39)	t = -0.18 d = 0.01
Peer problems									
Teacher-report	1-3	1.22 (0.34)	1.20 (0.33)	1.24 (0.35)	t = 1.44 d = 0.12	1.36 (0.50)	1.35 (0.51)	1.36 (0.48)	t = 0.25 d = 0.02
Child-report	T1: 1-2 T2: 1-4	1.18 (0.19)	1.16 (0.19)	1.20 (0.20)	t = 2.51* d = 0.21	2.17 (0.33)	2.17 (0.33)	2.16 (0.33)	t = -0.37 d = 0.03
Parent-report	1-3	1.26 (0.35)	1.25 (0.34)	1.27 (0.36)	t = 0.70 d = 0.06	1.36 (0.44)	1.36 (0.44)	1.37 (0.45)	t = 0.28 d = 0.02
Conduct problems									
Parent-report	1-3	1.35 (0.37)	1.33 (0.37)	1.38 (0.37)	t = 1.65 d = 0.14	1.38 (0.39)	1.37 (0.40)	1.38 (0.38)	t = 0.31 d = 0.03

Notes. [a] Not assessed at T1; n_{girls} = 304, n_{boys} = 295; *df* for all independent *t*-tests was 597.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Table I-9. Stabilities, correlations with age, and intercorrelations of the validation constructs

	Aggr-T T1	Peer-T T1	Peer-C T1	Peer-P T1	Cond-P T1	Aggr-T T2	Aggr-C T2	Aggr-S T2	Peer-T T2	Peer-C T2	Peer-P T2	Cond-P T2
Age T1	.00	.12**	.06	.01	.02	.03	.13**	-.09*	.11**	.02	.05	.04
Aggr-T T1		.38***	.19***	.20***	.33***	**.45***	.29***	.07	.33***	.23***	.28***	.32***
Peer-T T1			.26***	.31***	.26***	.22***	.21***	.05	**.31***	.26***	.35***	.30***
Peer-C T1				.29***	.21***	.17***	.12**	.03	.27***	**.29***	.32***	.25***
Peer-P T1					.52***	.31***	.21***	.15***	.35***	.29***	**.64***	.43***
Cond-P T1						.35***	.22***	.15***	.38***	.26***	.43***	**.61***
Aggr-T T2							.37***	.15***	.50***	.19***	.40***	.44***
Aggr-C T2								.22***	.25***	.18***	.35***	.30***
Aggr-S T2									.09*	.11**	.14***	.15***
Peer-T T2										.45***	.58***	.48***
Peer-C T2											.46***	.34***
Peer-P T2												.58***
Cond-P T2												

Notes. Aggr = aggressive behavior; Peer = reer problems; Cond = conduct Problems; T = teacher-report; C = child-report; P = parent-report; stabilities of constructs are presented in bold.

* $p < .05$. ** $p < .01$. *** $p < .001$.

5.2 Longitudinal Reciprocity between Theory of Mind and Aggression in Middle Childhood (Paper 2)

Abstract

Theory of mind is one of the most important cognitive factors in social information-processing, and deficits in theory of mind have been linked to aggressive behavior in childhood. The present longitudinal study investigated reciprocal links between theory of mind and two forms of aggression – physical and relational – in middle childhood with three data waves over 3 years. Theory of mind was assessed by participants' responses to cartoons, and physical and relational aggression were assessed through teacher reports in a community sample of 1657 children (mean age at Time 1: 8 years). Structural equation modeling analyses showed that theory of mind was a negative predictor of subsequent physical and relational aggression, both from Time 1 to Time 2 as well as from Time 2 to Time 3. Moreover, relational aggression was a negative predictor of theory of mind from Time 1 to Time 2. There were no significant gender or age differences in the tested pathways. The results suggest that reciprocal and negative longitudinal relations exist between children's theory of mind and aggressive behavior. Our study extends current knowledge about the development of such relations across middle childhood.

Keywords: aggressive behavior, theory of mind, middle childhood, longitudinal study, relational aggression, physical aggression.

Holl, A. K., Kirsch, F., Rohlf, H., Krahé, B., & Elsner, B. (2018). Longitudinal reciprocity between theory of mind and aggression in middle childhood. *International Journal of Behavioral Development*, 42, 257-266. doi:10.1177/0165025417727875

Aggressive behavior is one of the most pressing social problems in our world today, because of its detrimental effects on the lives and health of human beings (Dahlberg & Krug, 2002). Already in childhood, aggression has negative consequences, for instance, in the form of maladjustment of both children who show aggressive behavior (e.g. Crick et al., 2006) and children who are the targets of that behavior (e.g. Crick et al., 2001). Social cognition in general and theory of mind (ToM) in particular have been recognized as core processes involved in the development of aggressive behavior (Huesmann, 1998). However, whether ToM is negatively or positively linked to aggression and whether ToM is a cause or even a consequence of aggression are issues that have not yet been sufficiently addressed. Thus, the present study aimed at examining potential reciprocal relations between ToM and aggressive behavior in middle childhood.

Aggressive behavior is defined as "any form of behavior directed toward the goal of harming or injuring another live being who is motivated to avoid such treatment" (Baron & Richardson, 1994, p. 7). Harming others can be done by physical means, for example by hitting or kicking, known as *physical aggression* (PHY-A; e.g. Crick, Casas, & Ku, 1999), or by damaging their friendships or feelings of inclusion in the peer group, for example by spreading rumors, known as *relational aggression* (REL-A; e.g. Crick & Grotpeter, 1995). Some authors use the terms indirect or social aggression instead of REL-A. However, all three constructs essentially refer to the same form of aggression (Archer & Coyne, 2005), we refer to REL-A, even when the cited authors used an alternative term. Regarding the developmental course of the forms of aggression, the majority of children, girls in particular, reduce their use of PHY-A (Côté, Vaillancourt, Barker, Nagin, & Tremblay, 2007) and tend to increase their use of REL-A (Crick et al., 1999) during middle childhood.

There are typical gender differences in the two forms of aggression, with boys engaging in PHY-A more often than girls (e.g. Lansford et al., 2012). Theoretical explanations for this gender difference in PHY-A include hormonal, evolutionary, and social role accounts (see Krahé, 2013, for review). With regard to REL-A, several studies have found that girls engage in REL-A more often than boys (e.g. Ostrov, Murray-Close, Godleski, & Hart, 2013). One explanation is that girls tend to have more dyadic relationships than boys, which seems to fuel REL-A (Murray-Close, Ostrov, & Crick, 2007). However, other studies revealed either no gender differences in REL-A (e.g. Lansford et al., 2012) or higher scores for boys (e.g. Henington, Hughes, Cavell, & Thompson, 1998). These mixed findings highlight the importance of conducting further research on the developmental courses of PHY-A and REL-

A during middle childhood, thereby including both forms of aggression and differentiating developmental patterns for girls and boys.

The Social Information-Processing (SIP) Theory of Aggression

Crick and Dodge's (1994) social information-processing (SIP) model proposes that deficits and biases in processing social information, especially the misinterpretation or neglect of important social cues, can cause social maladjustment in children, including aggressive behavior. SIP is conceptualized as a sequence of six processing stages: (1) encoding and (2) interpretation of social cues, (3) clarification of individual goals, (4) access to or construction of potential behavioral responses, (5) decision for a response, and (6) behavioral enactment. This stepwise approach has helped to identify differences in SIP patterns in children who show higher as compared to lower levels of aggression. For instance, aggressive children tend to attribute hostile intentions to their peers in ambiguous social situations more often than non-aggressive children (de Castro, Veerman, Koops, Bosch, & Monshouwer, 2002, SIP Steps 1 and 2). Furthermore, aggressive children generate more aggressive responses than their non-aggressive peers (de Castro, Merk, Koops, Veerman, & Bosch, 2005, SIP Step 4).

In terms of the chronification of aggressive behavior, Crick and Dodge (1994) assumed that the relation between social experiences and SIP is reciprocal. They proposed that mental representations of social experiences are stored in long-term memory, in a so-called database, and are then accessed in each of the postulated SIP steps. Thus, social knowledge is retrieved and, in turn, shapes behavior at each step. Consequently, social cognition influences the occurrence of aggressive behavior, and, in turn, memories of past aggressive behavior influence social cognition. In one of the few prospective studies that have examined this reciprocal link, Lansford, Malone, Dodge, Pettit, and Bates (2010) demonstrated a dynamic developmental cascade in which SIP and aggressive behavior influenced one another over the course of middle childhood. Crick and Dodge (1994) proposed investigating the impact of social experiences, such as aggressive behavior, on social cognitions. This could foster understanding of possible mechanisms by which children develop outcome expectations for specific social situations, which, in turn, children may use to reach a response decision (SIP Step 5).

ToM, which enables individuals to "predict what others are going to do on the basis of their desires and beliefs" (Frith & Frith, 2010, p. 165), is involved in at least some steps of the SIP model (Harvey, Fletcher, & French, 2001). ToM is typically distinguished into two facets:

Making inferences regarding others' beliefs, intentions, or desires, refers to *cognitive* ToM, and inferring others' emotions refers to *affective* ToM (e.g. Shamay-Tsoory, Harari, Aharon-Peretz, & Levkovitz, 2010). In the last decade, behavioral (Devine & Hughes, 2013) as well as neuroscientific studies (Vetter, Weigelt, Döhnel, Smolka, & Kliegel, 2014) confirmed the further development of ToM after the accomplishment of understanding of false belief during preschool age. The development of cognitive ToM (Devine & Hughes, 2013) as well as the understanding of advanced affective ToM (Vetter et al., 2014) are assumed to continue throughout middle childhood and adolescence.

Both facets of ToM are assumed to influence at least some of the steps postulated in the SIP model. For instance, ToM plays an important role in attributional processes and in forming a mental representation of a social situation (Teufel, Fletcher, & Davis, 2010), both of which are part of encoding and interpretation of social cues (SIP Steps 1 and 2). Furthermore, ToM seems to play a role in reaching a response decision (SIP Step 5), as has been shown in a peer-coordination game, in which 6-year-old children adjusted their response decisions by applying their ToM skills (Grueneisen, Wyman, & Tomasello, 2015).

Theory of Mind and Aggressive Behavior

Numerous studies have examined the relation between ToM and aggressive behavior in children. However, most of these studies yielded mixed results, were situated in preschool age, were cross-sectional, or did not control for earlier levels of behavior. Some studies supported the "social skills deficit view" described above, which assumes that deficits in ToM co-occur with a pattern of SIP that fosters aggressive behavior (Crick & Dodge, 1994). For instance, deficits in ToM were correlated with behavioral problems in 2-year-olds (Hughes & Ensor, 2006), as well as in a clinical sample of 5- to 9-year-olds (Fahie & Symons, 2003). Similarly, preschool-aged children with superior ToM were rated as being less aggressive toward their peers (Capage & Watson, 2001).

However, in contrast to the social skills deficit view, Sutton, Smith, and Swettenham (1999a) proposed that children who show aggressive behavior toward their peers, instead of being "socially inadequate", are more socially competent than their victims. This reasoning comes primarily from research on bullying, as a specific form of aggressive behavior that is ongoing and involves a power asymmetry between perpetrator and victim (Olweus, 2013). In children with low ToM abilities, bullying may result from deficits in SIP (e.g. misinterpretation

of social cues), as described above. However, bullies also may have superior ToM abilities that enable them to manipulate others, which was demonstrated particularly for ringleader-bullies in middle childhood (Sutton, Smith, & Swettenham, 1999b). Sutton et al. (1999a) further argue that ToM abilities may be particularly relevant for the use of relational forms of bullying, because children with higher ToM abilities are better able to anticipate the victim's reactions to REL-A, and by which means they can best harm their victim. Although this assumption was formulated for relational bullying, it may be applicable to REL-A in general, because most relationally aggressive acts require the involvement of others. In line with this reasoning, in one of the few longitudinal studies addressing the link between ToM and aggression, 5-year-old children's ToM skills were positively related to REL-A one year later, but only in children with average or low levels of prosocial behavior (Renouf et al., 2010). In this study, PHY-A was unrelated to ToM.

Because most previous studies were cross-sectional (Gomez-Garibello & Talwar, 2015), or did not control for earlier levels of behavior, not much is known about the direction of effects between ToM and different forms of aggression, PHY-A and REL-A, particularly in middle childhood. There is some evidence that ToM influences later aggression, for instance in preschool age (Renouf et al., 2010). However, the reverse path is also possible, as has been found in preschoolers whose aggressive behavior at 2.5 years negatively predicted their ToM skills 4 months later (Song, Volling, Lane, & Wellman, 2016). The SIP model proposes reciprocal relations between social cognition (i.e. ToM) and social behavior, based on a continuous retrieval of stored mental representations of social experiences from one's database (Crick & Dodge, 1994). In addition, not much is known about potential changes in this link during middle childhood, when the further development of ToM abilities (e.g. Hughes & Devine, 2015) coincides with changes in PHY-A and REL-A (e.g. Côté et al., 2007). After the transition to school, a variety of social experiences, such as involvement in aggressive behavior, could influence the further development of ToM (see Hughes & Leekam, 2004, for a review on the links between ToM and social relationships). Yet, to our knowledge there are no longitudinal studies that examined both the potential path from ToM to PHY-A and REL-A, and the reverse path from these aggression forms to ToM at the same time over the course of middle childhood.

Additionally, the roles of gender and age as potential moderators of the links between ToM and aggression require further study. As explained above, girls and boys differ in their expression of REL-A and PHY-A (e.g. Lansford et al., 2012), therefore the links to ToM could also differ between genders. Until now, the very few studies that considered gender differences

in the relation between ToM and aggression produced inconsistent results (e.g. Kokkinos, Voulgaridou, Mandrali, & Parousidou, 2016; Renouf et al., 2010). With regard to age, only a few studies have considered age as a moderator in the relation between ToM and aggression. As this link was mainly examined in children of preschool age, even less is known about possible age differences in middle childhood. One study found a moderation by age in the relation between REL-A and ToM in South-American children (Gomez-Garibello & Talwar, 2015).

The Current Study

In our three-wave longitudinal study, we examined the reciprocal relations of ToM with PHY-A and REL-A in middle childhood in a large community sample over a period of 3 years. At the first data wave (T1), participants had a mean age of 8 years, the second wave (T2) was conducted about 9 months later, and the third wave (T3) another 24 months later. Using structural equation modeling, we investigated the reciprocal relations while controlling for earlier levels of ToM and partialling out stable between-person differences in aggressive behavior, respectively.

On the basis of the SIP model (Crick & Dodge, 1994) and in line with earlier empirical evidence (e.g. Capage & Watson, 2001), we expected a negative path from ToM to subsequent PHY-A. Regarding the path from ToM to later REL-A, the SIP model also suggests a negative path, whereas other research, particularly in the bullying tradition, suggests a positive path (e.g. Renouf et al., 2010). We examined these competing predictions in our analyses. Additionally, in line with the SIP approach, we postulated prospective paths from aggressive behavior to ToM. Past social experiences are stored in long-term memory and should be retrieved in every step of SIP. Since ToM is involved in at least some of the SIP steps, it might be influenced by earlier aggression. We therefore hypothesized negative paths from both PHY-A and REL-A to subsequent ToM.

In addition, we explored possible gender and age differences in the reciprocal associations between aggression and ToM. This was deemed important, because the inconsistent findings across previous studies point toward the need for further investigations to clarify the potential moderating influence of gender and age on the paths between ToM and aggression. To examine whether the proposed relations could be confirmed for all subsamples

of our study, we conducted separate multi-group analyses with the gender groups and with three age-groups, respectively.

Method

Participants

At the first data wave (T1) 1657 children (52.1% girls) between 6 and 11 years (M = 8.36, SD = 0.95, range 6.23–11.33) participated. At the second wave (T2), about 9 months after T1, 1611 children (51.8% girls) participated again (M = 9.12, SD = 0.93, range 7.12–11.90). At the final wave (T3), about 24 months after T2, 1501 children (51.7% girls) participated again (M = 11.07, SD = 0.92, range 9.12–13.76). All children who took part at T1 were included in our analyses, and missing data due to dropout or incomplete measures were handled by the full information maximum likelihood procedure (FIML; Enders & Bandalos, 2001), resulting in a sample size of N = 1657 children. The children who no longer participated at T3 had significantly higher PHY-A and REL-A, and lower ToM and verbal ability scores at T1 than those children who completed T3.

Children were recruited at 33 community primary schools representing a variety of rural and urban areas. About 33.7% of the mothers and 36.9% of the fathers reported holding a university degree, 21.6% and 13.5% a university entrance qualification, 42.9% and 48.0% a vocational level qualification, and 1.8% and 1.7% no or a low level of school graduation.

Material and Procedure

The study was part of a larger research project on intrapersonal developmental risk factors in childhood and adolescence from a longitudinal perspective. Data were collected in individual sessions at the children's schools in a private room by a trained research assistant. Class teachers received individual questionnaires for each participating child. All procedures were approved by the Ethics Committee at the University of Potsdam and by the Ministry of Education, Youth and Sport of the Federal State of Brandenburg, Germany. For each child, informed consent was obtained from the primary caregiver, and the children gave verbal consent.

Theory of mind. At T1 and T2, cartoons were used to assess ToM (Sebastian et al., 2012). Based on a pilot study, 12 cartoons (six for cognitive, six for affective ToM) of the original 20 cartoons were chosen as representing an average difficulty level for the studied age

group. The cartoons were presented to the children on netbooks using E-Prime 2.0 Professional (Psychology Software Tools, 2012). In each trial, the first three pictures depicted a little story with two characters (A and B). In the affective stories, character A displayed an emotion, for example, sadness after losing a boat on a river (Figure II-1); in the cognitive stories, character A displayed an intention or desire, for instance, wanting to pick apples from a tree. Then, two further pictures were presented simultaneously, each displaying a possible ending. The "correct" ending implied that character B understood A's mental state (e.g. by comforting or helping); in the "incorrect" ending, character B showed uncaring behavior. The positions of the types of endings were varied across trials, and children were asked to select the picture they thought represented the correct ending by pressing the F-key or the J-key (on a QWERTZ keyboard) for the left or right picture, respectively. Children received feedback only in two initial exercise trials, but not when they worked through the 12 trials. The trials were presented in random order.

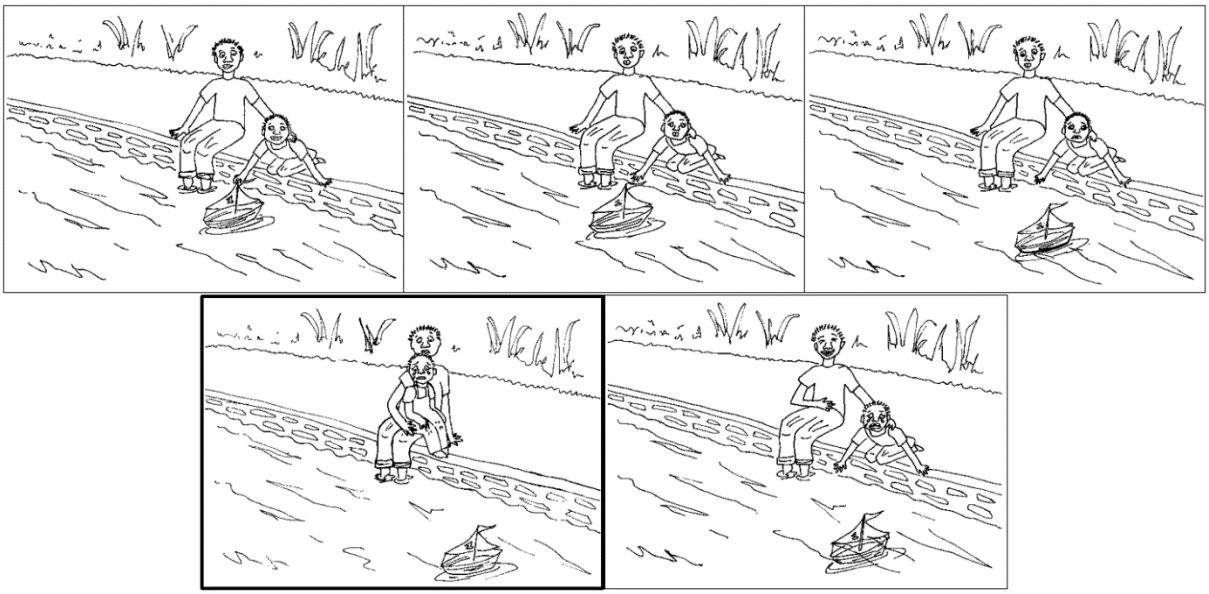

Figure II-1. Example cartoon story 'Boat' for assessing affective ToM

Notes. The first three pictures were shown consecutively; after that, two pictures displaying possible endings appeared simultaneously (here, the correct answer has a bold frame).

The number of correct responses were summed up separately, resulting in one score for affective ToM and one for cognitive ToM, both with a range of 0–6. In the analyses, we included ToM as a latent factor with the affective and the cognitive ToM score as indicators. Ordinal alpha, which is appropriate for binary items (Gadermann, Guhn, & Zumbo, 2012),

indicated good internal consistency for our ToM measure at both time points (α_{T1} = .81, α_{T2} = .82).

Aggressive behavior. At all three time points, aggressive behavior was assessed using six items adapted from the Children's Social Behavior Scale-Teacher Form (CSBS-T; Crick, 1996). Classroom teachers indicated how often during the last 6 months the child had shown behaviors of PHY-A (three items, e.g. "hit, shove, or push peers") and REL-A (three items, e.g. "try to exclude certain peers from peer group activities"), using a response scale ranging from 1 (*never*), 2 (*once a month or rarely*), 3 (*several times a month*), 4 (*several times a week*), and 5 (*every day*). In the hypothesis-testing analysis, we used manifest mean scores of PHY-A and REL-A, and included random intercepts for both forms, as described below. Internal consistency was excellent (PHY-A: α_{T1} = .93, α_{T2} = .94, α_{T3} = .93; REL-A: α_{T1} = .91, α_{T2} = .92, α_{T3} = .90).

Verbal ability. Because children's ToM skills can be confounded with their verbal ability (Milligan, Astington, & Dack, 2007), it is essential to control for the latter. At T1, we used a vocabulary test from the Potsdam-Illinois Test for Psycholinguistic Abilities (P-ITPA; Esser, Wyschkon, Ballaschk, & Hänsch, 2010) to assess children's word knowledge and ability to detect relations between words. Scores were normed for children's age.

Statistical Analyses

Structural equation models were conducted with Mplus (Version 7.4; Muthén & Muthén, 1998–2015), and descriptive analyses were computed with SPSS (Version 23). In all structural equation models, the robust maximum likelihood estimator (MLR) was used to account for deviations from normality of the data. As recommended by Hamaker, Kuiper, and Grasman (2015) we included a latent random intercept for both forms of aggressive behavior to control for stable between-person differences. This procedure provides a more accurate examination of within-person changes. As ToM was measured only at two time points, we were not able to include a random intercept for this construct (random intercepts require at least three time points). Instead, we included ToM as a latent factor with the affective and the cognitive ToM scores as indicators. Metric measurement invariance over time was established for ToM.

Age and gender were covariates for all constructs, and verbal ability assessed at T1 was an additional covariate for ToM. In addition, separate multi-group analyses were conducted to examine potential gender and age differences. A model in which all paths were restricted to be

equal between the groups (fully constrained) was compared with a model in which all paths were freed between the groups (fully unconstrained). Model comparisons were based on χ^2 differences used with scale corrections as proposed by Satorra and Bentler (2001) due to the application of the MLR estimator. Age-groups ("younger" vs. "middle" vs. "older") were defined based on splitting the sample by age at T1 at the 33th and 66th percentile ($Md_{33\%}$ = 7.83 years, $Md_{66\%}$ = 8.86 years).

The percentage of missing values ranged from 1.6% to 33.3% (for the sample size of each measure see Table II-1). The rate of missing data was highest on the teacher reports of aggression, due to not-returned questionnaires. The FIML procedure was implemented in all analyses to handle missing data (Enders & Bandalos, 2001).

Overall model fit was evaluated according to Hu and Bentler (1999), with RMSEA ≤ .06, CFI ≥ .95, TLI ≥ .95, and SRMR ≤ .08 indicating a good fit. The χ^2-statistic is reported to provide complete model fit information, but it was not evaluated as an absolute index of model fit because it is biased for large samples (Schermelleh-Engel, Moosbrugger, & Müller, 2003). Effect sizes for *t*-tests are reported as Cohen's *d* with the conventional interpretation of *d* = .20 as small, *d* = .50 as medium, and *d* = .80 as large effects (Cohen, 1988). For our structural equation model, we report standardized regression coefficients that can be interpreted in a similar way as Cohen's *d*.

Results

Descriptive Statistics, Gender Differences, and Correlations

Table II-1 presents the descriptive statistics and the results of statistical tests for gender differences on the aggressive behavior measures as well as the ToM scores. Teacher-rated PHY-A scores were significantly higher for boys than for girls at each data wave, with medium to large effect sizes (*d*s ≥ 0.66). For REL-A, the gender difference was small (*d*s ≤ 0.14), but boys were rated as significantly more relationally aggressive than girls at T2 and T3. With regard to ToM, no significant gender differences emerged.

PHY-A scores decreased significantly from T1 to T2, $t(1117) = -2.68$, $p < .01$, $d = .06$, but were stable between T2 and T3, $t(843) = -0.99$, $p = .32$, $d = .03$. REL-A scores did not vary significantly between T1 and T2, $t(1115) = 1.86$, $p = .06$, $d = .05$, or between T2 and T3, $t(840) = -1.66$, $p = .10$, $d = .06$. ToM scores increased significantly from T1 to T2, $t(1543) = 16.22$, $p < .001$, $d = .48$.

Table II-1. Descriptive statistics and gender differences

	n (*n girls*)	Range	Total *M* (*SD*)	Girls *M* (*SD*)	Boys *M* (*SD*)	Gender difference
Physical Aggression[a]						
T1	1411 (732)	1 - 5	1.49 (0.79)	1.22 (0.52)	1.79 (0.91)	$t(1063.36) = -14.52^{***}$ $d = 0.77$
T2	1148 (585)	1 - 5	1.48 (0.77)	1.18 (0.46)	1.79 (0.90)	$t(825.45) = -14.23^{***}$ $d = 0.85$
T3	1107 (551)	1 - 5	1.44 (0.75)	1.20 (0.49)	1.67 (0.88)	$t(864.37) = -11.07^{***}$ $d = 0.66$
Relational Aggression[a]						
T1	1408 (732)	1 - 5	1.50 (0.70)	1.47 (0.70)	1.54 (0.70)	$t(1406) = -1.65$ $d = 0.10$
T2	1147 (584)	1 - 5	1.56 (0.76)	1.51 (0.73)	1.60 (0.79)	$t(1145) = -2.15^{*}$ $d = 0.12$
T3	1105 (550)	1 - 5	1.50 (0.71)	1.45 (0.66)	1.55 (0.76)	$t(1086.01) = -2.22^{*}$ $d = 0.14$
Theory of Mind[b]						
T1	1630 (848)	0 - 12	10.09 (1.90)	10.17 (1.86)	10.00 (1.95)	$t(1628) = 1.83$ $d = 0.09$
T2	1567 (800)	0 - 12	10.91 (1.44)	10.92 (1.46)	10.89 (1.42)	$t(1565) = 0.41$ $d = 0.02$

Notes. [a] mean score; [b] sum score. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

Bivariate correlations between all variables are displayed in Table II-2. Age was positively correlated with ToM at T1 and T2, but uncorrelated with aggressive behavior. All correlations between ToM and aggressive behavior were negative. Stability of ToM between T1 and T2 was relatively low, whereas PHY-A and REL-A showed moderate to high stabilities between the time points. PHY-A and REL-A were strongly positively associated with each other at all time points.

Table II-2. Intercorrelations of study variables

	T1 Verbal Ability	T1 Theory of Mind	T1 Physical Aggression	T1 Relational Aggression	T2 Theory of Mind	T2 Physical Aggression	T2 Relational Aggression	T3 Physical Aggression	T3 Relational Aggression
T1 Age	-.11***	.16***	-.03	-.05	.15***	-.03	-.04	-.02	-.02
T1 Verbal Ability		.15***	-.09**	-.09***	.17***	-.10**	-.11***	-.13***	-.14***
T1 Theory of Mind			-.12***	-.05	*.33***	-.11***	-.08**	-.11***	-.12***
T1 Physical Aggression				.62***	-.07**	*.72***	.45***	*.48***	.32***
T1 Relational Aggression					-.09***	.45***	*.53***	.33***	*.36***
T2 Theory of Mind						-.08**	-.08**	-.08**	-.08*
T2 Physical Aggression							.66***	*.50***	.37***
T2 Relational Aggression								.31***	*.40***
T3 Physical Aggression									.66***
T3 Relational Aggression									

Notes. Stabilities are highlighted in italics; $N = 1657$. * $p < .05$. ** $p < .01$. *** $p < .001$.

Hypothesis Testing

To test our first set of hypotheses, we computed a structural equation model with latent factors for ToM at T1 and T2, and manifest mean scores for PHY-A and REL-A at T1, T2, and T3, including random intercepts for both REL-A and PHY-A. In doing so, we investigated the longitudinal reciprocal relations between ToM and PHY-A and REL-A, respectively, controlling for earlier levels of ToM and for stable between-person differences in aggressive behavior. Furthermore, all constructs were controlled for age and gender, and ToM was additionally controlled for T1 verbal ability. Metric measurement invariance over time was established in a step-up approach for PHY-A, REL-A, and ToM from T1 to T2. The overall fit of our final model was good, RMSEA = .03, 90% CI [.02, .04]; CFI = .99; TLI = .98; SRMR = .03; $\chi^2(33) = 71.38$, $p < .001$. The resulting model is presented in Figure II-2.

Reciprocal effects of ToM, REL-A, and PHY-A. In line with our hypotheses, we found significant negative paths from ToM at T1 to PHY-A at T2 ($\beta = -.08$, $p < .05$), and from ToM at T2 to PHY-A at T3 ($\beta = -.12$, $p < .01$). Similarly, the hypothesized paths from ToM at T1 to REL-A at T2 ($\beta = -.09$, $p < .05$) as well as from ToM at T2 to REL-A at T3 ($\beta = -.11$, $p < .05$) were significant and negative. The reverse negative path was confirmed only from REL-A at T1 to ToM at T2 ($\beta = -.12$, $p < .05$), but not from PHY-A at T1 to ToM at T2 ($\beta = .09$, $p = .13$). In addition to the expected paths, a negative path from PHY-A at T2 to REL-A at T3 was found ($\beta = -.16$, $p < .01$).

Gender differences. To examine gender differences, we compared the fully constrained multi-group model by gender with the fully unconstrained model (fit constrained model: RMSEA = .05, 90% CI [.04, .05]; CFI = .95; TLI = .93; SRMR = .06; $\chi^2(95) = 256.03$, $p < .001$; fit unconstrained model: RMSEA = .03, 90% CI [.03, .04]; CFI = .98; TLI = .96; SRMR = .04; $\chi^2(65) = 127.14$, $p < .001$). The difference between these two models was significant, $\Delta\chi^2(30) = 119.43$, $p < .001$. In a next step, we computed a revised and final model, in which we constrained all paths to be equal, except the paths that were found to differ significantly between boys and girls in the fully unconstrained model. This revised model showed a good model fit, RMSEA = .03, 90% CI [.02, .04]; CFI = .98; TLI = .97; SRMR = .04; $\chi^2(86) = 153.49$, $p < .001$, which was significantly better than that of the fully constrained model, $\Delta\chi^2(9) = 79.87$, $p < .001$, and not significantly worse than that of the fully unconstrained model, $\Delta\chi^2(21) = 27.83$, $p = .145$. Based on the revised model, there were no significant gender differences in the hypothesized paths ($ps \geq .12$). The only significant gender differences were found in the concurrent correlations between PHY-A and REL-A, which were stronger for boys (rs ranging

from .54 to .69, *p*s < .001) than for girls (*r*s ranging from .42 to .45, *p*s < .001). In addition, on a within-person level, REL-A at T1 predicted REL-A at T2 only in girls (β = .26, *p* < .001), but not in boys (β = .09, *p* = .17).

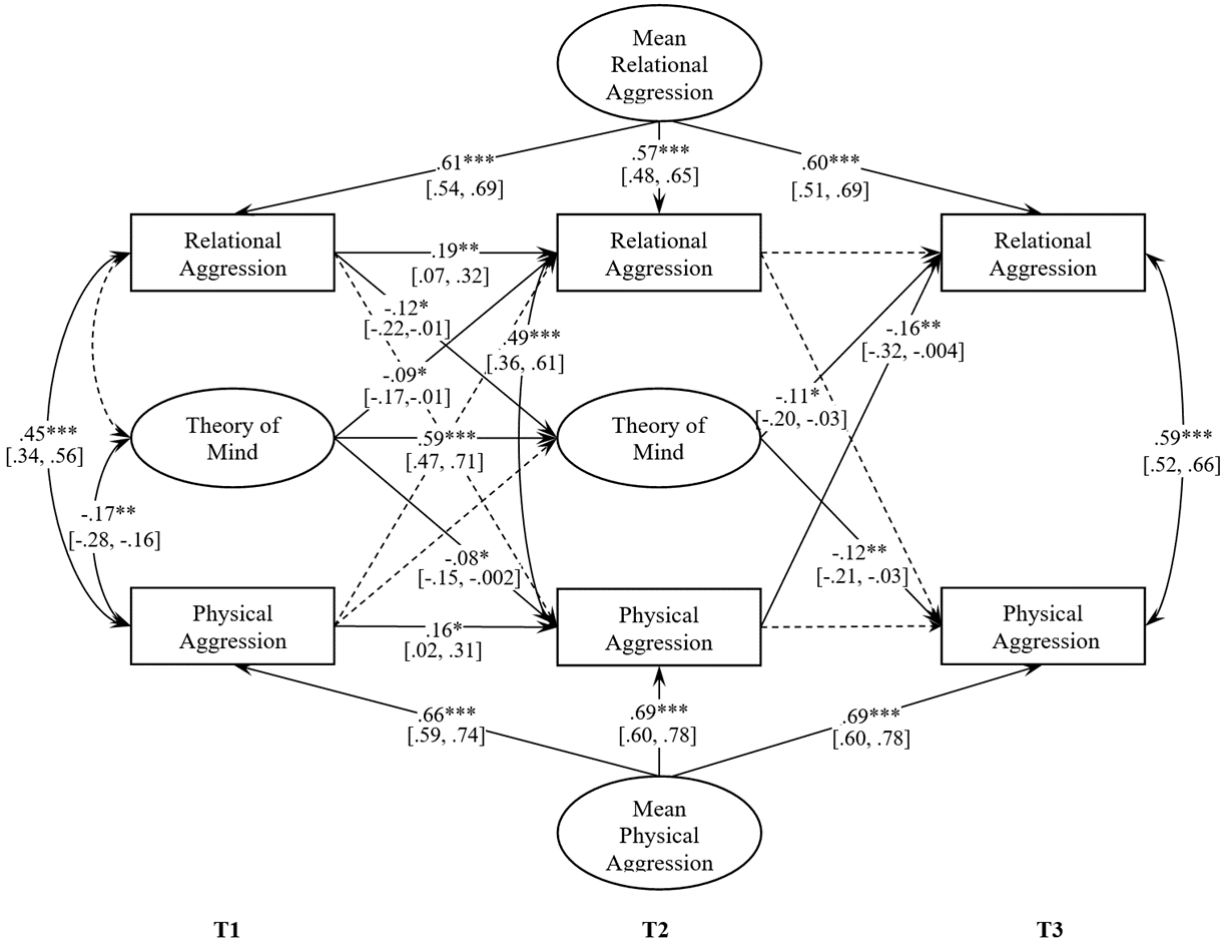

Figure II-2. Structural equation model for physical aggression, relational aggression and theory of mind from T1 to T3

Notes. Standardized coefficients are displayed; dashed lines display non-significant paths; number in square brackets display 95% CIs; mean relational aggression and mean physical aggression partial out between-person stability over time (random intercept); all variables were controlled for age and gender, theory of mind was additionally controlled for verbal ability; *N* = 1657; Model fit: RMSEA = .03, 90% CI [.02, .04]; CFI = .99; TLI = .98; SRMR = .03; $\chi^2(33) = 71.38$. * *p* < .05. ** *p* < .01. *** *p* < .001.

Age differences. To examine age differences, we followed the same procedure with the three age-groups; fit constrained model: RMSEA = .03, 90% CI [.02, .03]; CFI = .98; TLI = .98; SRMR = .06; $\chi^2(164) = 227.32$, *p* < .001; fit unconstrained model: RMSEA = .03, 90% CI [.02, .04]; CFI = .99; TLI = .97; SRMR = .05; $\chi^2(104) = 154.20$, *p* < .001. The fully constrained

model did not fit worse than the fully unconstrained model, $\Delta\chi^2(60) = 75.41$, $p = .09$, indicating that there was no significant moderation by age.

Discussion

The primary aim of this research was to study the reciprocal relations between ToM (assessed by children's responses to a cartoon task) and teacher-reported PHY-A and REL-A over time in a large community sample of girls and boys in middle childhood. Our longitudinal study with three waves of measurement covered a time period of about 3 years. Past studies on the association of ToM and aggression were mostly cross-sectional, were conducted at preschool age, and did not separate distinct forms of aggression. Therefore, we investigated the reciprocal relations between ToM and PHY-A and REL-A during middle childhood in a prospective design using structural equation modeling, controlling for age, gender, and verbal ability as well as earlier levels of ToM and partialling out stable between-person differences in aggressive behavior.

Consistent with our predictions, we found that ToM was a significant negative predictor of both PHY-A and REL-A. This was true for the paths from T1 to T2 as well as from T2 to T3. These results indicate that children with lower ToM scores were rated by their teachers to be more physically and relationally aggressive at a later occasion, while controlling for stable between-person differences in aggressive behavior. Even though we did not examine mediating processes in our study, these results are in line with the predictions derived from the SIP model (Crick & Dodge, 1994; Harvey et al., 2001) that deficits in ToM may lead to biased or deficient SIP, which in turn may lead to more aggressive behavior. For example, previous research revealed that the encoding and interpretation of situational cues were less accurate in children with early deficits in ToM (e.g. Choe, Lane, Grabell, & Olson, 2013). Our findings do not support the proposition by Sutton et al. (1999a) that children may use their proficient ToM to manipulate others, joining other studies that also failed to find differences in ToM between bullies and noninvolved children (e.g. Gini, 2006). However, the consideration of potential moderating variables, such as prosocial behavior (e.g. Renouf et al., 2010), may be important to identify those children that exploit their ToM skills in an aggressive fashion.

Although the effect sizes were small, the present findings are consistent with previous studies that examined relations between ToM and aggression (e.g. Kokkinos et al., 2016). Considering the complex nature of the emergence of aggressive behavior as postulated by the

SIP model (Crick & Dodge, 1994), many intra- and interpersonal factors can have an impact (e.g. see Krahé, 2013, for a review). Our study demonstrated ToM to be an important, but certainly not the only intrapersonal predictor for the development of aggressive behavior during middle childhood.

Turning to the postulated reverse paths from aggression to ToM, higher scores in REL-A, but not PHY-A, at T1 predicted lower ToM skills at T2. The negative path from REL-A to ToM is in line with the SIP model, postulating reciprocal paths between social cognition and aggressive behavior (Crick & Dodge, 1994). The reason for the missing path from PHY-A to ToM may be that the different functions of aggression in terms of proactive versus reactive aggression were not considered in our study. In past research, only reactive PHY-A was linked to low socio-cognitive abilities (Crick & Dodge, 1996). Proactive PHY-A, however, is an instrumental and planned behavior (Vitaro & Brendgen, 2005) and may be positively linked to ToM. These opposing relations could have cancelled each other out to eliminate the path from PHY-A to ToM. Nevertheless, it remains unclear why this would have been the case only for PHY-A and not for REL-A.

In addition to the hypothesized paths, the analyses revealed a significant negative path from PHY-A at T2 to REL-A at T3. This result indicates that physically aggressive children tend to reduce their relationally aggressive behavior at a later occasion. This is in contrast to previous longitudinal studies that have found a positive relation or no relation between these two forms of aggression over time (e.g. Card, Stucky, Sawalani, & Little, 2008). However, these studies usually examined aggression on a population level (e.g. by applying traditional cross-lagged panel models), whereas our study examined aggression on the level of the individual (by including random intercepts; Hamaker et al., 2015). The divergence in findings may be an expression of Simpson's paradox (Kievit, Frankenhuis, Waldorp, & Borsboom, 2013), that is, the direction of a link depends on the level of analysis (i.e. population vs. individual). To our knowledge, there are only a few studies to date that examined PHY-A and REL-A from a person-centered perspective. These studies examined trajectory groups of PHY-A and REL-A as well as the co-occurrence of these groups (e.g. Côté et al., 2007; Ettekal, & Ladd, 2015), and found substantial proportions of children who – on an individual level – showed different trajectories of PHY-A and REL-A (e.g. increasing in PHY-A, but decreasing in REL-A). Therefore, future studies should further investigate the interplay of PHY-A and REL-A longitudinally on an intrapersonal as well as on an interpersonal level.

In further analyses, we explored potential age and gender differences in the reciprocal relations between ToM, PHY-A, and REL-A in middle childhood. We did not find any differences in the reciprocal relations of ToM and the forms of aggression between girls and boys, and between age-groups. However, we found that the correlation between REL-A and PHY-A was higher for boys than for girls, which is consistent with previous research (Card et al., 2008). In addition, we discovered significant gender differences in the mean level of aggressive behavior. At each wave, boys were rated by their teachers to show significantly more PHY-A than girls. This difference was of medium to large size and is in accordance with earlier studies (e.g. Lansford et al., 2012). Similarly, boys were rated to show more REL-A than girls at T2 and T3. This difference was small in size and is in contrast to other studies that found girls as compared to boys to exhibit more REL-A (e.g. Card et al., 2008; Ostrov et al., 2013); albeit, some studies also found no gender differences in REL-A (e.g. Lansford et al., 2012), or higher scores for boys (e.g. Henington et al., 1998). The present results may have been due to characteristics of our male or female subsample or to population differences, which we could not examine in our study. A more fine-grained analysis of gender differences in REL-A remains an important topic for future research. Regarding age, we did not find differences in the reciprocal relations between ToM and the two forms of aggression across the age-groups of our sample. There were no correlations of age and PHY-A or REL-A, but there were positive correlations of age and ToM. In line with previous research (e.g. Devine & Hughes, 2013), this indicates that ToM is still improving with increasing age in middle childhood. Altogether, we conclude that apart from gender differences in the absolute level of aggressive behavior, in the concurrent correlations between REL-A and PHY-A, and bivariate correlations between age and ToM skills, the prospective relations between aggression and ToM are not moderated by gender or age.

In conclusion, the present findings provided the first longitudinal evidence on reciprocal relations between ToM and PHY-A, as well as REL-A, as distinct forms of aggression, in middle childhood. Our results of negative reciprocal relations between ToM and REL-A or PHY-A, respectively, are consistent with the social skills deficit view of Crick and Dodge (1994), and they do not support the proposition by Sutton et al. (1999a) that aggressive children have advanced ToM abilities. This suggests that improving ToM abilities may reduce rather than promote PHY-A and REL-A in middle childhood. The impact of REL-A on children's well-being is usually evaluated by teachers and parents as being less severe than that of PHY-A (e.g. Hurd & Gettinger, 2011). In contrast, our study showed that trivialization of REL-A may have detrimental effects on socio-cognitive development. Therefore, interventions that aim

at reducing REL-A and/or fostering ToM would be indicated for interrupting the vicious circle of deficient socio-cognitive abilities promoting aggressive behavior and vice versa. Following our results, gender differences do not seem to play a prominent role in the examined processes. Therefore, interventions should focus equally on girls and boys.

Strengths and Limitations

We believe our study has several strengths. It is based on a large sample of children attending community elementary schools, and it includes three data waves covering middle childhood over a total of about 3 years. We controlled for stable between-person differences in aggression and for past levels of ToM (construct stability), fostering a causal interpretation of the paths (Marmor & Montemayor, 1977). Further, we distinguished between two forms of aggression, PHY-A and REL-A, which have been found in some previous studies to differ between girls and boys (e.g. Lansford et al., 2012), and we explored potential gender and age differences. Finally, as one of the first studies, we investigated not only the path from ToM to the two forms of aggression, but also the reverse paths. Crick and Dodge (1994) already emphasized the importance of studying the effect of social experiences, such as aggressive behavior, on social cognitions. This can promote understanding of how children develop outcome expectations for social situations, which they then apply in reaching response decisions.

At the same time, there are some limitations to our study. First, we used the prominent SIP model by Crick and Dodge (1994) as a framework for our study, but we did not include measures specifically designed for SIP, for example video vignettes (Lansford et al., 2006). Consequently, the specific underlying mechanisms by which ToM affects SIP cannot be inferred. Further, we did not examine other mediating mechanisms, such as empathy or moral disengagement (e.g. Kokkinos et al., 2016). Future studies should therefore integrate measures for specific cognitive processes that are assumed to influence aggressive behavior.

Second, the time intervals differed between T1 and T2 (about 9 months), and T2 and T3 (about 24 months). Consequently, prospective effects were compared between unequal time periods; however, it is even more remarkable that ToM predicted PHY-A and REL-A at both time intervals. Another limitation is that our ToM measure appeared somewhat too easy for our sample, especially at T2. These ceiling effects reduced the variance in the ToM measures, making it harder to detect potential effects and to identify children that are very skilled in ToM. Future studies should continue to develop time- and resource-efficient ToM measures that

create sufficient variability in middle childhood. Finally, the use of teacher reports to assess aggressive behavior can be considered as a limitation of our study. Teachers may not be fully aware of the extent to which children engage in aggressive behavior, in particular when it comes to REL-A, which includes more covert behavior (e.g. spreading rumors).

Despite these limitations, our findings highlight the consideration of reciprocal relations between ToM and aggressive behavior over the course of middle childhood. Future research should further investigate how aggressive behavior and ToM develop over the whole course of childhood and how this feeds back into the development of the respective other part over time. In the long run, this information can help to reduce the detrimental effects of aggressive behavior on children's lives and health.

References

Archer, J., & Coyne, S. (2005). An integrated review of indirect, relational, and social aggression. *Personality and Social Psychology Review*, *9*, 212-230.

Baron, R. A., & Richardson, D. S. (1994). *Human aggression* (2nd ed.). New York, NY: Plenum Press.

Capage, L., & Watson, A. C. (2001). Individual differences in theory of mind, aggressive behavior, and social skills in young children. *Early Education & Development*, *12*, 613–628.

Card, N., Stucky, B., Sawalani, G., & Little, T. (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.

Choe, D. E., Lane, J. D., Grabell, A. S., & Olson, S. L. (2013). Developmental precursors of young school-age children's hostile attribution bias. *Developmental Psychology*, *49*, 2245–2256.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Côté, S. M., Vaillancourt, T., Barker, E. D., Nagin, D., & Tremblay, R. E. (2007). The joint development of physical and indirect aggression: Predictors of continuity and change during childhood. *Development and Psychopathology*, *19*, 37–55.

Crick, N. R. (1996). The role of overt aggression, relational aggression, and prosocial behavior in the prediction of children's future social adjustment. *Child Development*, *67*, 2317–2327.

Crick, N. R., Casas, J. F., & Ku, H.-C. (1999). Relational and physical forms of peer victimization in preschool. *Developmental Psychology*, *35*, 376–385.

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.

Crick, N. R., & Dodge, K. A. (1996). Social information-processing mechanisms in reactive and proactive aggression. *Child Development*, *67*, 993–1002.

Crick, N. R., & Grotpeter, J. K. (1995). Relational aggression, gender, and social-psychological adjustment. *Child Development*, *66*, 710–722.

Crick, N. R., Nelson, D. A., Morales, J. R., Cullerton-Sen, C., Casas, J. F., & Hickman, S. E. (2001). Relational victimization in childhood and adolescence: I hurt you through the grapevine. In J. Juvonen & S. Graham (Eds.), *Peer harassment in school: The plight of the vulnerable and victimized* (pp. 196–214). New York, NY: Guilford Press.

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.

Dahlberg, L. L., & Krug, E. G. (2002). Violence - a global public health problem. In E. G. Krug, L. L. Dahlberg, J. A. Mercy, A. B. Zwi, & R. Lozano (Eds.), *World report on violence and health* (pp. 3–21). Geneva, Switzerland: World Health Organisation.

de Castro, B. O., Merk, W., Koops, W., Veerman, J. W., & Bosch, J. D. (2005). Emotions in social information processing and their relations with reactive and proactive aggression in referred aggressive boys. *Journal of Clinical Child and Adolescent Psychology*, *34*, 105–116.

de Castro, B. O., Veerman, J. W., Koops, W., Bosch, J. D., & Monshouwer, H. J. (2002). Hostile attribution of intent and aggressive behavior: A meta-analysis. *Child Development*, *73*, 916–934.

Devine, R. T., & Hughes, C. (2013). Silent films and strange stories: Theory of mind, gender, and social experiences in middle childhood. *Child Development*, *84*, 989–1003.

Enders, C., & Bandalos, D. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling, 8*, 430–457.

Esser, G., Wyschkon, A., Ballaschk, K., & Hänsch, S. (2010). *P-ITPA: Potsdam-Illinois Test für Psycholinguistische Fähigkeiten* [P-ITPA: Potsdam-Illinois test for psycholinguistic abilities]. Göttingen, Germany: Hogrefe.

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.

Fahie, C. M., & Symons, D. K. (2003). Executive functioning and theory of mind in children clinically referred for attention and behavior problems. *Journal of Applied Developmental Psychology*, *24*, 51–73.

Frith, U., & Frith, C. (2010). The social brain: Allowing humans to boldly go where no other species has been. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *365*, 165–176.

Gadermann, A. M., Guhn, M., & Zumbo, B. D. (2012). Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. *Practical Assessment, Research & Evaluation*, *17*, 1–13.

Gini, G. (2006). Social cognition and moral cognition in bullying: what's wrong? *Aggressive Behavior*, *32*, 528-539.

Gomez-Garibello, C., & Talwar, V. (2015). Can you read my mind? Age as a moderator in the relationship between theory of mind and relational aggression. *International Journal of Behavioral Development*, *39*, 552–559.

Grueneisen, S., Wyman, E., & Tomasello, M. (2015). "I know you don't know I know..." Children use second-order false-belief reasoning for peer coordination. *Child Development*, *86*, 287–293.

Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. P. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, *20*, 102–116.

Harvey, R. J., Fletcher, J., & French, D. J. (2001). Social reasoning: An influence on aggression. *Clinical Psychology Review*, *21*, 459–469.

Henington, C., Hughes, J. N., Cavell, T. A., & Thompson, B. (1998). The role of relational aggression in identifying aggressive boys and girls. *Journal of School Psychology, 36*, 457–477.

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

Huesmann, R. L. (1998). The role of social information processing and cognitive schema in the acquisition and maintenance of habitual aggressive behavior. In R. G. Geen & E. Donnerstein (Eds.), *Human aggression. Theories, research, and implications for social policy* (pp. 73–109). San Diego, CA: Academic Press.

Hughes, C., & Devine, R. T. (2015). Individual differences in theory of mind from preschool to adolescence: Achievements and directions. *Child Development Perspectives, 9*, 149-153.

Hughes, C., & Ensor, R. (2006). Behavioral problems in 2-year-olds: Links with individual differences in theory of mind, executive function and harsh parenting. *Journal of Child Psychology and Psychiatry*, *47*, 488–497.

Hughes, C., & Leekam, S. (2004). What are the links between theory of mind and social relations? Review, reflections and new directions for studies of typical and atypical development. *Social Development*, *13*, 590–619.

Hurd, H. D., & Gettinger, M. (2011). Mothers' and teachers' perceptions of relational and physical aggression in pre-school children. *Early Child Development and Care*, *181*, 1343–1359.

Kievit, R. A., Frankenhuis, W. E., Waldorp, L. J., & Borsboom, D. (2013). Simpson's paradox in psychological science: A practical guide. *Frontiers in Psychology*, *4*, Article 513.

Kokkinos, C. M., Voulgaridou, I., Mandrali, M., & Parousidou, C. (2016). Interactive links between relational aggression, theory of mind, and moral disengagement among early adolescents. *Psychology in the Schools*, *53*, 253–269.

Krahé, B. (2013). *The social psychology of aggression* (2nd ed.). Hove, England: Psychology Press.

Lansford, J. E., Malone, P. S., Dodge, K. A., Crozier, J. C., Pettit, G. S., & Bates, J. E. (2006). A 12-year prospective study of patterns of social information processing problems and externalizing behaviors. *Journal of Abnormal Child Psychology*, *34*, 715–724.

Lansford, J. E., Malone, P. S., Dodge, K. A., Pettit, G. S., & Bates, J. E. (2010). Developmental cascades of peer rejection, social information processing biases, and aggression during middle childhood. *Development and Psychopathology*, *22*, 593–602.

Lansford, J. E., Skinner, A. T., Sorbring, E., Di Giunta, L., Deater-Deckard, K., Dodge, K. A., … Chang, L. (2012). Boys' and girls' relational and physical aggression in nine countries. *Aggressive Behavior*, *38*, 298–308.

Marmor, G. S., & Montemayor, R. (1977). The cross-lagged panel design: A review. *Perceptual and Motor Skills, 45*, 883–893.

Milligan, K., Astington, J. W., & Dack, L. A. (2007). Language and theory of mind: Meta-analysis of the relation between language ability and false-belief understanding. *Child Development*, *78*, 622–646.

Murray-Close, D., Ostrov, J. M., & Crick, N. R. (2007). A short-term longitudinal study of growth of relational aggression during middle childhood: Associations with gender, friendship intimacy, and internalizing problems. *Development and Psychopathology*, *19*, 187–203.

Muthén, L. K., & Muthén, B. O. (1998-2015). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Olweus, D. (2013). School bullying: Development and some important challenges. *Annual Review of Clinical Psychology*, *9*, 751–780.

Ostrov, J. M., Murray-Close, D., Godleski, S. A., & Hart, E. J. (2013). Prospective associations between forms and functions of aggression and social and affective processes during early childhood. *Journal of Experimental Child Psychology*, *116*, 19-36.

Psychology Software Tools. (2012). E-Prime 2.0 [Computer Software]. Retrieved from http://www.pstnet.com

Renouf, A., Brendgen, M., Parent, S., Vitaro, F., Zelazo, P. D., Boivin, M., ... Séguin, J. R. (2010). Relations between theory of mind and indirect and physical aggression in kindergarten: Evidence of the moderating role of prosocial behaviors. *Social Development*, *19*, 535–555.

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, *66*, 507–514.

Schermelleh-Engel, H., Moosbrugger, K., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, *8*, 23-74.

Sebastian, C. L., Fontaine, N. M. G., Bird, G., Blakemore, S.-J., De Brito, S. A., McCrory, E. J. P., & Viding, E. (2012). Neural processing associated with cognitive and affective theory of mind in adolescents and adults. *Social Cognitive and Affective Neuroscience*, *7*, 53–63.

Shamay-Tsoory, S. G., Harari, H., Aharon-Peretz, J., & Levkovitz, Y. (2010). The role of the orbitofrontal cortex in affective theory of mind deficits in criminal offenders with psychopathic tendencies. *Cortex*, *46*, 668–677.

Song, J.-H., Volling, B. L., Lane, J. D., & Wellman, H. M. (2016). Aggression, sibling antagonism, and theory of mind during the first year of siblinghood: A developmental cascade model. *Child Development*, *87*, 1250–1263.

Sutton, J., Smith, P. K., & Swettenham, J. (1999a). Bullying and "theory of mind": A critique of the "social skills deficit" view of anti-social behavior. *Social Development*, *8*, 117–127.

Sutton, J., Smith, P. K., & Swettenham, J. (1999b). Social cognition and bullying: Social inadequacy or skilled manipulation? *British Journal of Developmental Psychology*, *17*, 435-450.

Teufel, C., Fletcher, P. C., & Davis, G. (2010). Seeing other minds: Attributed mental states influence perception. *Trends in Cognitive Sciences*, *14*, 376–382.

Vetter, N., Weigelt, S., Döhnel, K., Smolka, M. N., Kliegel, M. (2014). Ongoing neural development of affective theory of mind in adolescence. *Social Cognitive and Affective Neuroscience*, *9*, 1022–1029.

Vitaro, F., & Brendgen, M. (2005). Proactive and reactive aggression: A developmental perspective. In R. E. Tremblay, W. W. Hartup, & J. Archer (Eds.), *Developmental origins of aggression* (pp. 178–201). New York, NY: Guilford Press.

5.3 Physical Attractiveness, Peer Problems and Aggressive Behavior in Childhood: A Three-Year Longitudinal Study (Paper 3)

Abstract

Children's physical attractiveness is examined as a distal factor in the development of aggressive behavior. Based on socialization theory, it is predicted that less attractive individuals experience more peer problems), which in turn can foster aggression. These relations were tested in a three-year longitudinal study covering three waves. The sample consisted of 1,043 children (Time1: 6-10 years). Physical attractiveness was assessed by independent raters at T1 and T3. At all time points, aggression was measured by teacher-reports, and peer problems by self-reports. Results supported socialization theory by revealing an indirect path from attractiveness at T1 and aggression at T3.

Keywords: physical attractiveness, peer problems, aggression, childhood, longitudinal methodology.

Kirsch, F., Busching, R., Rohlf, H., & Krahé, B. (under review). Physical attractiveness, peer problems and aggressive behavior in childhood: A three-year longitudinal study.

Physical appearance plays an important role in many areas of personal and professional life. One of the most relevant aspects of appearance is physical attractiveness, that is "the extent to which a stimulus person is pleasing to observe" (Patzer, 2006, p. 324). In the last five decades, psychology has studied the determinants and effects of physical attractiveness intensively. One robust finding is the observation that an individual's physical attractiveness is strongly correlated with others' perception and expectations about that person, indicating the existence of a physical attractiveness stereotype (Dion, Berscheid, & Walster, 1972). According to this stereotype, attractive individuals, are viewed as more positive compared to their less attractive counterparts on a broad range of personality traits, including social competence and adjustment ("What is beautiful is good"; for a meta-analytic review, see Eagly, Ashmore, Makhijani, & Longo, 1991). Conversely, unattractive individuals are viewed as having a more negative personality, including less sociability and more trait aggression ("What is ugly is bad"; e.g., Griffin & Langlois, 2006; Dion, 1973).

As a result of these differences in social perception, attractive individuals are in general treated more favorably by their social environment (for a meta-analytic review, see Langlois et al., 2000). For example, experimental research has revealed that attractive compared to unattractive individuals received more lenient sanctions for crimes, misconduct or low work performance (e.g., Berkowitz & Frodi, 1979; Mazzella & Feingold, 1994; Commisso & Finkelstein, 2012). Also, other people try harder to get in touch with more attractive individuals (e.g., Gueguen, Lourel, Charron, Fischer-Lokou, & Lamy, 2009). In the long run, such differences in the social treatment by others can have a considerable impact on an individual's development. In a longitudinal study by Benzeval, Green, and Macintyre (2013), physical attractiveness by the age of 15 years was a positive predictor for a range of positive life outcomes 20 years later, including occupational and social success.

Among all the advantages, being physically attractive has a proximal and powerful impact on social life from childhood onward. Attractive children have more friends and are more popular in their peer group than less attractive children (for a meta-analytic review, see Langlois et al., 2000). Conversely, unattractive children experience more problems with their peers than their more attractive counterparts (e.g., Sweeting & West, 2010).

These appearance-dependent differences in social preference can have a huge impact on an individual's health, well-being, and behavior (e.g., Goswami, 2012; Umberson, Chen, House, Hopkins, & Slaten, 1996; Umberson & Montez, 2010). Low physical attractiveness may be a distal risk factor for maladjustment by undermining the quality of an individual's social

relationships. Rosen, Underwood, and Beron (2011) showed that low attractiveness in early adolescence indirectly predicted internalizing problems through higher peer victimization, which in turn led to the development of internalizing symptoms. A similar mechanism can be expected for other problematic outcomes. Peer problems are not only associated with internalizing problems but have also been established as a major risk factor for externalizing problems, in particular aggression (for a review, see Leary, Twenge, & Quinlivan, 2006). Thus, lower attractiveness may be associated with the development of aggressive behavior through the underlying impact of problematic peer interactions. However, to our knowledge, this assumption has not been addressed so far. Thus, the aim of the present study was to examine the link between attractiveness and aggression including peer problems as a possible mechanism underlying this relation.

Physical Attractiveness and Aggressive Behavior

For the purpose of our study, aggressive behavior is defined as behavior with the intention to harm (Baron & Richardson, 1994). Most studies that considered physical attractiveness and aggressive behavior examined the interaction between these two constructs to predict various outcomes (e.g., popularity, Rosen & Underwood, 2010). Only few studies investigated attractiveness as a predictor of aggressive behavior. In the study by Bobadilla, Metze, and Taylor (2013), participants were given an opportunity to aggress against a provocateur, before (unprovoked aggression) and after (reactive aggression) a provocation in the laboratory. The authors found a significant negative correlation of participants' attractiveness and their aggressive behavior, but only in males. The less attractive male students were found to exhibit more unprovoked and reactive aggressive behavior in comparison to average and highly attractive students, supporting the "what is ugly is bad" stereotype.

In a study by Langlois and Downs (1979), children at the ages of three and five years were observed during dyadic free play, and their aggressive behavior was coded. While there was no association between physical attractiveness and aggression in three-year-old children, unattractive compared to attractive boys among the five-year-olds showed more aggressive behavior towards their play partner. Unattractive girls only showed more aggression when their play partner was also unattractive. Additionally, in a study by Serketich and Dumas (1997) and a follow-up study by Dumas, Nilsen, and Lynch (2001), it was reported that children who were referred for their aggressive behavior by their teachers were rated as less physically attractive by independent college students than non-aggressive children. These studies provide some

evidence that less attractive individuals exhibit more aggressive behavior than their more attractive peers.

As discussed above, one important mechanism for the negative correlation between attractiveness and aggression may be problems in peer relationships, which is consistent with socialization theories (e.g., Langlois et al., 2000). In the case of childhood aggression, the causal chain may look like this: A child with a low level of physical attractiveness activates the "what is ugly is bad" stereotype, which includes the expectation that this child is more likely to show aggression (e.g., Dion, 1973). Others who generate these negative expectations in turn treat this unattractive child in a more unfavorable way (e.g., reject or avoid the child; Zakin, 1983). The resulting peer problems trigger aggressive behavior in the long run (e.g., Dodge et al., 2003; Rohlf, Busching, & Krahé, in press). Over time, the child that is expected to act more aggressively (because of his or her appearance) might develop into a more aggressive person, resulting in a self-fulfilling prophecy. Conversely, for more attractive children, the activated attractiveness stereotype leads others to form more positive expectations of that child ("what is beautiful is good"). This includes the expectation of high social competence (e.g., Langlois et al., 2000) and low antisocial behavior (e.g., Dion, 1973). Consequently, the more attractive children are more popular among their peers (e.g., Langlois et al., 2000) and receive more favorable treatment (e.g., Smith, 1985), giving them the opportunity to develop into a socially competent, non-aggressive person. Conversely, the unattractive children may miss these opportunities for social practice. In this sense, physical attractiveness acts as a distal developmental factor as it determines the socialization to which an individual is exposed. High physical attractiveness can function as a social asset, whereas low attractiveness functions as a social hindrance.

In contrast to the findings and theoretical considerations reviewed above, there is some evidence for a positive link between physical attractiveness and aggressive behavior. Borch, Hyde, and Cillessen (2011) found that attractive high school students are more likely to exhibit overt and relational aggressive behavior. The authors argue that more attractive individuals can afford to show aggressive behavior because their attractiveness protects them from social sanctions, so they exploit their social asset. Given these findings in adolescents, our study aimed to test whether younger children's physical attractiveness can act as a distal factor acting toward or against the development of aggressive behavior.

The Present Study

Based on the evidence discussed above, our study was designed to examine the impact of physical attractiveness on the development of aggressive behavior. We assumed that there is a negative relation between physical attractiveness and aggressive behavior that can be explained by the different consequences appearance can have on peer relationships. Although there is some evidence for an "attractive is more aggressive" perspective (Borch et al., 2011), we expect the opposite pattern ("attractive is less aggressive") since we focus on the development of aggressive behavior over time in which potential socialization processes towards the attractiveness stereotype may unfold a cumulative detrimental effect. Our study examined the prospective relations between physical attractiveness, peer problems, and aggressive behavior in a longitudinal study with a large sample of children in middle childhood. The study included three data waves covering three years. At the first and third measurement point (T1 and T3), physical attractiveness was rated by independent adult raters based on photos of the children's faces. Peer problems were assessed by self-reports, and aggressive behavior was assessed by teacher reports at all time points.

We examined the following hypotheses:

Hypothesis 1: Children's physical attractiveness is a negative prospective predictor of peer problems. Children who are rated as less attractive are more likely to report problems with their peers at subsequent data waves than their more attractive peers.

Hypothesis 2: Children's peer problems prospectively predict aggressive behavior. The more problems children have with their peers, the more aggressively they behave over a subsequent period of time.

Hypothesis 3: Physical attractiveness indirectly affects aggressive behavior through peer problems. Less attractive children show an increase of aggressive behavior over time as a result of experiencing more problems in their peer relationships.

In addition, we examined potential differences in our proposed associations in relation to sex and to the distinction between physical and relational aggression. There is evidence that the proposed paths from low attractiveness to peer problems (Hypothesis 1) and from peer problems to aggression (Hypothesis 2) are similar for boys and girls (e.g., Sweeting, & West, 2010; Dodge et al., 2003). Therefore, we checked a potential moderation by sex. In addition, we examined whether the paths would differ for physical and relational aggression. Previous research suggest that peer relationships should be predictive for both forms (e.g., Zimmer-

Gembeck, Geiger, & Crick, 2005), but it is unclear whether the indirect paths from low attractiveness via peer problems might differ. Finally, by assessing physical attractiveness at two time points in our longitudinal study, we were able to assess the developmental stability of physical attractiveness over the course of middle childhood, a period largely neglected in past research.

Method

Participants and Procedure

The participants were part of a larger sample of 1,657 children in Germany who took part in a longitudinal study on intrapersonal developmental risk factors in childhood and adolescence. Informed consent for children's participation in the study was obtained from their parents, and passive consent was obtained from the children themselves. Children whose parents did not give permission to record visual material of the children (n = 474) and those for whom the recorded visual material was low in quality (n = 140) were excluded, yielding a final sample at T1 of N = 1,043 children (50.4% female). Most children (87.9%) fell into the age range of 7 to 9 years (M = 8.33, SD = 0.94; overall range: 6 to 10 years). Children included in the sample did not differ from those who were excluded on the measures of peer problems and aggressive behavior (ts $\leq$ |0.96|, ps $\geq$ .34). At T2, which took place about 9 months after T1 (M = 9.14 months, SD = 1.80), 1,016 children participated again. At T3, which took place about 2 years after T2 (M = 23.83 months, SD = 1.66), the remaining sample consisted of 977 children.

In total, only 6.3% participants at T1 did not participate at the final time point T3 (dropouts). Dropouts differed from the remaining participants at T1 by higher scores in overall aggressive behavior (t[53.14] = -2.20, p = .03, d = 0.36) and physical aggression (t[53.56] = -2.21, p = .03, d = 0.37). To handle the resulting missing values we used the Full Information Maximum Likelihood (FIML) procedure in all analyses (Enders & Bandalos, 2001). Data collection was conducted by trained project staff at the participants' school at all three time points. The instruments were approved by the Ethics Committee of the authors' university and the Ministry of Education, Youth and Sport of the Federal State of Brandenburg, Germany.

Physical Attractiveness

Ratings of physical attractiveness were obtained at T1 and T3 based on square photos (240×240 px) of children's faces (facial attractiveness). The photos were extracted from existing video material used for a behavioral observation task, as described in Rohlf and Krahé (2015). To extract the photos from the videos, we used the screenshot tool PicPick (Version 4.1.2; NGWIN, 2016). We extracted the photos in a way that the children's faces were roughly the same size within the photos and clothes were only minimally visible. Criteria for photo extraction were a frontal view on the child's face, a neutral facial expression, visible eyes, and a sufficient photo quality. To further improve standardization between the photos and to reduce the influence of the background within a photo, a black mask was added using Adobe Photoshop CS2 (Version 9.0; see Figure III-1).

Figure III-1. Example of a masked photo

Note. The photos were not pixelated in the rating procedure.

Each photo was rated for physical attractiveness on a response scale from 1 (*not at all attractive*) to 7 (*very attractive*). T1 ratings were done by 24 student raters (22 female; Age *M* = 25.38 years, *SD* = 2.96), T3 ratings were done by 23 student raters (21 female; Age *M* = 25.87 years, *SD* = 2.85). Previous research revealed that individuals of different ages share a consensus about who is considered to be attractive and who is not (e.g., Slater et al., 1998; Ma, Xu, & Luo, 2015), and obtaining ratings of children's attractiveness from independent adult raters is a commonly used procedure (e.g., Rosen & Underwood, 2010). The ratings took place online using the survey application LimeSurvey (Version 2.05). Each rater completed three rating sessions. Each session took place on a separate day and contained a third of the photos in randomized order. The sessions themselves also were run through in randomized order. Inter-

rater agreement was excellent, as indicated by high intra-class coefficients at both time points (ICCs = .95). For each child, attractiveness scores for T1 and T3 were calculated by averaging all corresponding ratings.

Peer Problems

Peer problems were assessed using children's self-reports. At T1 and T2, we used eight items that capture aspects of problematic peer relationships particularly relevant in this age group: five items from the Social Integration subscale of the Questionnaire on Social and Emotional Experiences at School of Elementary School Children (FEESS; Rauer & Schuck, 2003, 2004; example item: "The other children often laugh at me") and three items from the Peer Acceptance subscale of the German version of the Harter-Scales (Asendorpf & van Aken, 1993; example item: "I am often asked if I want to play", reverse coding). The response options were 1 (no) and 2 (yes). Due to the binary nature of the response scale, we calculated ordinal alpha instead of Cronbach's alpha as a measure of reliability (Gadermann, Guhn, & Zumbo, 2012). Ordinal Alpha was good for both time points (αt1 = .83, αt2 = .87). As the children were substantially older at the final time point, the measure was adjusted to remain age appropriate and to still capture relevant aspects of peer problems. Thus, at T3, we used seven items of the FEESS (including the five items presented at T1 and T2) and a more elaborated 4-point response scale with 1 (no), 2 (rather no), 3 (rather yes), and 4 (yes). Cronbach's alpha was good (αt3 = .82). For all time points, a sum score was computed as a measure of self-reported peer problems.

Aggressive Behavior

Children's aggressive behavior was operationalized by teacher reports at all three time points using six items that were based on the Children´s Social Behavior Scale - Teacher Form (CSBS-T; Crick, 1996). Teachers were asked to indicate the frequency of different acts of aggression in the last six months on a scale from 1 (*never*) to 5 (*daily*). Three items referred to physically aggressive behavior (example item: "How often did this child hit, shove, or push peers?") and three considered relationally aggressive behavior (example item: "How often did this child try to exclude certain peers from peer group activities?"). All items were averaged to form a global measure of aggressive behavior with high internal consistency (α_{t1} = .91, α_{t2} = .93, α_{t3} = .93). Additionally, to examine potential differences in the two forms of aggressive behavior, we

computed mean scores for physical (α_{t1} = .93, α_{t2} = .94, α_{t3} = .93) as well as relational aggression (α_{t1} = .90, α_{t2} = .92, α_{t3} = .90).

Data Analysis Plan

SPSS (Version 23) was used for descriptive computations. The hypotheses were tested through path analysis using Mplus (Version 7.4; Muthén & Muthén, 1998-2015). In all of our path analyses, the robust maximum likelihood (MLR) estimator was used to handle non-normality of our data. Due to the trait-like nature of aggressive behavior, we included a random intercept for this measure, as recommended by Hamaker, Kuiper, and Grasman (2015). The random intercept partials out time-invariant stability (i.e., trait components) of individual differences in aggressive behavior. Thus, this procedure provides a separation of the within-person change over time from the between-person differences in the level of a construct. Compared to traditional analytical approaches, this approach facilitates a more accurate examination of within-person processes in our dependent variable. The assessments of peer problems differed between the time points (i.e., in items and response format), and physical attractiveness was only measured at two waves. As this procedure requires invariant assessments across at least three time points, we did not include random intercepts for these two constructs.

Indirect effects were computed using R (Version 3.3.1) following a Monte Carlo procedure with 20,000 repetitions (Hayes & Scharkow, 2013). Model fit was evaluated based on the recommendations of Hu & Bentler (1999): CFI should be > .95, RMSEA should be < .06, and SRMR should be < .08. For the sake of completeness, χ^2 is reported but was not used in the evaluation of model fit due to its sensitivity for large sample sizes (Schermelleh-Engel, Moosbrugger, & Müller, 2003). To check for sex differences, we conducted χ^2 difference tests. Specifically, we compared a model in which all paths were restricted to be equal between girls and boys (fully constrained) with a second model in which all paths were allowed to vary between the sexes (fully unconstrained).

Results

Descriptive Statistics, Sex Differences and Correlations

Table III-1 presents descriptive statistics and information about sex differences. Girls were rated as significantly more physically attractive than boys at both T1 and T3 (*t*s ≥ 3.80, *p*s <

.001, *d*s ≥ 0.24). Boys were rated to be more overall aggressive at all time points (*t*s ≥ |4.69|, *p*s < .001, *d*s ≥ 0.36). The sex difference was significant for physical aggression at all time point (*t*s ≥ |8.12|, *p*s < .001, *d*s ≥ 0.60), and for relational aggression at T2 only (*t*[719] = -2.14, *p* = .03, *d* = 0.16). For peer problems, boys reported more peer problems than did girls at T1 and T2 (*t*s ≥ |1.97|, *p*s < .05, *d*s ≥ 0.12). Due to the sex differences in our examined variables, we included sex as a covariate in our path analyses.

Table III-1. Descriptive statistics and sex differences

	n (*n* girls)	Range	Total *M* (*SD*)	Girls *M* (*SD*)	Boys *M* (*SD*)
Attractiveness[a]					
T1	1043 (526)	1-7	3.91 (0.84)	*4.01* (0.91)	*3.81* (0.75)
T3	877 (435)	1-7	3.91 (0.85)	*4.02* (0.91)	*3.80* (0.77)
Peer problems[b]					
T1	1034 (522)	8-16	9.44 (1.63)	*9.30 (1.60)*	*9.59 (1.64)*
T2	1002 (504)	8-16	9.41 (1.65)	*9.30 (1.55)*	*9.51 (1.74)*
T3	961 (483)	7-28	10.55 (3.56)	10.50 (3.58)	10.61 (3.53)
Total aggression[a]					
T1	889 (450)	1-5	1.50 (0.67)	*1.36* (0.56)	*1.65* (0.73)
T2	720 (356)	1-5	1.52 (0.72)	*1.34* (0.53)	*1.69* (0.83)
T3	691 (344)	1-5	1.45 (0.67)	*1.34* (0.53)	*1.57* (0.77)
Physical aggression[a]					
T1	896 (452)	1-5	1.49 (0.79)	*1.23* (0.54)	*1.76* (0.90)
T2	727 (360)	1-5	1.47 (0.78)	*1.18* (0.46)	*1.75* (0.91)
T3	717 (350)	1-5	1.44 (0.75)	*1.22* (0.51)	*1.65* (0.88)
Relational aggression[a]					
T1	891 (450)	1-5	1.51 (0.71)	1.49 (0.71)	1.54 (0.70)
T2	721 (356)	1-5	1.56 (0.78)	*1.50* (0.71)	*1.62* (0.84)
T3	695 (347)	1-5	1.49 (0.71)	1.46 (0.66)	1.51 (0.76)

Notes. Scores in italics differ significantly between girls and boys (*p* < .05); [a] mean score; [b] sum score.

Table III-2 presents manifest stability information, bivariate correlations of the examined constructs, and their correlations with age. Stabilities of peer problems and aggression were

moderate with *r*s ranging from .31 to .68. Physical attractiveness showed a substantial stability over three years with $r = .52$ ($p < .001$), indicating that attractiveness can be seen as a trait-like variable in middle childhood. Age was negatively correlated with T1 physical attractiveness and T3 peer problems. Consequently, we included age as a covariate in our path analyses. Aggressive behavior (total score and both forms) and peer problems were positively correlated with each other, indicating that more aggressive children also face more problems with their peers. An exception was the nonsignificant cross-sectional correlation at T3 between relational aggression and peer problems. Finally, as expected, physical attractiveness at both time points was exclusively negatively correlated with physical aggression and peer problems, both cross-sectionally and longitudinally. Thus, less attractive children were more physically aggressive and experienced more peer problems. Relational aggression was uncorrelated with physical attractiveness at T1 and T3.

Hypothesis Testing

To test the hypothesized model, a path analysis was computed using the total aggression scores, as displayed in Figure III-2. As prior multi-group analyses revealed no significant moderation by sex ($\Delta\chi^2[28] = 40.818$, $p = .06$), sex was included as a covariate in the single-group model. In addition, the model was statistically controlled for age and included a random intercept for aggressive behavior. Model fit was sufficient ($\chi^2[7] = 46.022$, $p < .001$; CFI = .970; SRMR = .025; RMSEA = .073).

Hypothesis 1 predicted a negative prospective link from physical attractiveness to peer problems. Controlling for peer problems at T1, attractiveness assessed at T1 negatively predicted peer problems at T2 ($\beta = -.06$, 95% CI [-.109, -.004], $p = .04$): Children, who were less attractive at T1 reported more problems with their peers about nine months later (T2). This result is consistent with our first hypothesis. The path from attractiveness at T1 to peer problems at T3 was nonsignificant ($\beta = -.01$, 95% CI [-.071, .045], $p = .65$).

Hypothesis 2 proposed a positive link from peer problems to aggressive behavior. Controlling for stable individual differences in aggression, peer problems at T2 positively predicted aggressive behavior at T3 ($\beta = .19$, 95% CI [.101, .270], $p < .001$). A significant path was also found from T1 peer problems to T2 aggression ($\beta = .13$, 95% CI [.051, .199], $p < .001$). Thus, children who experienced more problems with their peers subsequently exhibited more aggressive behavior. These results confirmed our second hypothesis.

Table III-2. Stabilities, correlations with age, and intercorrelations

	Attract T1	Peer Prob T1	Total Aggr T1	Phy Aggr T1	Rel Aggr T1	Peer Prob T2	Total Aggr T2	Phy Aggr T2	Rel Aggr T2	Attract T3	Peer Prob T3	Total Aggr T3	Phy Aggr T3	Rel Aggr T3
Age T1	-.08*	.04	-.03	-.02	-.03	.01	-.04	-.04	-.03	.02	-.10**	.01	-.01	.00
Attract T1		-.09**	-.08*	-.10**	-.05	-.10**	-.08*	-.10**	-.05	**.52*****	-.04	-.11**	-.14***	-.07
Peer Prob T1			.19***	.19***	.14***	**.52*****	.21***	.22***	.17***	-.15***	**.31*****	.18***	.20***	.11**
Total Aggr T1				.91***	.88***	.16***	**.64*****	.62***	.56***	-.12**	.24***	**.43*****	.45***	.36***
Phy Aggr T1					.60***	.17***	.63***	**.68*****	.48***	-.12**	.21***	.40***	**.45*****	.30***
Rel Aggr T1						.11**	.51***	.41***	**.53*****	-.09*	.20***	.36***	.34***	**.34*****
Peer Prob T2							.24***	.25***	.20***	-.17***	**.38*****	.19***	.22***	.12**
Total Aggr T2								.93***	.92***	-.03	.16***	**.45*****	.47***	.39***
Phy Aggr T2									.71***	-.05	.16***	.44***	**.50*****	.34***
Rel Aggr T2										-.01	.14***	.38***	.36***	**.38*****
Attract T3											-.12***	-.10*	-.13**	-.04
Peer Prob T3												.11**	.15***	.05
Total Aggr T3													.93***	.92***
Phy Aggr T3														.71***
Rel Aggr T3														

Notes. Stabilities are highlighted in bold; Attract = physical attractiveness, Peer Prob = peer problems, Total Aggr = total aggression score, Phy Aggr = physical aggression, Rel Aggr = relational aggression.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Hypothesis 3 proposed an indirect path from physical attractiveness to aggressive behavior through peer problems. The indirect effect was of small size, but significant with β = -.01 (95% CI [-.023, -.001]). Thus, children, who were rated less attractive at T1 showed more aggressive behavior at T3 with the link being partly attributable to more peer problems in the interim period (T2). This result is consistent with our third hypothesis. It is noteworthy that the direct paths from physical attractiveness at T1 to aggressive behavior at T2 (β = -.03, 95% CI [-.086, .027], p = .31) and T3 (β = -.07, 95% CI [-.144, .005], p = .07) were nonsignificant.

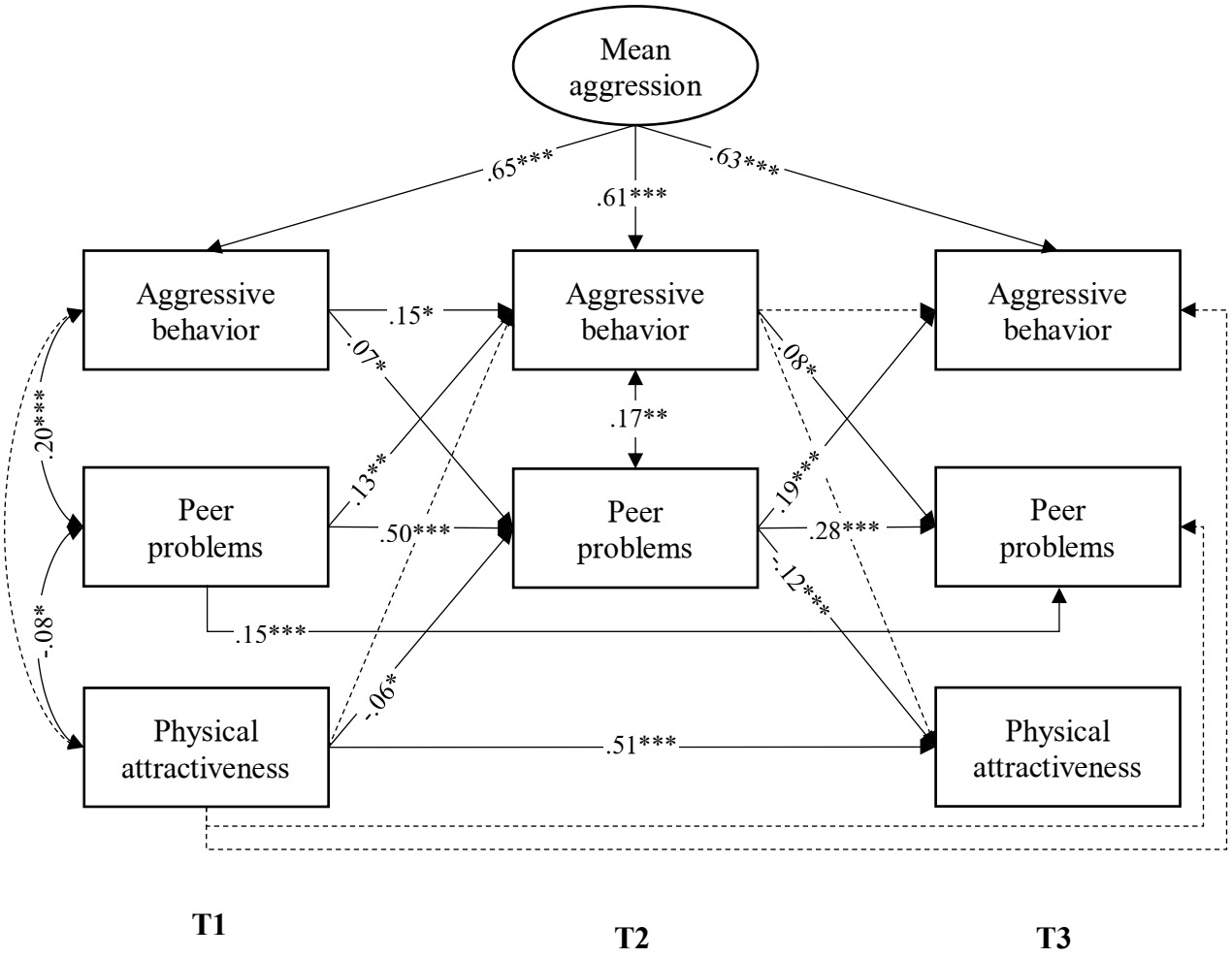

Figure III-2. Path analysis for physical attractiveness, peer problems and aggressive behavior from T1 to T3

Notes. Standardized coefficients are displayed; dashed lines display nonsignificant paths; mean aggression partials out between-person stability over time (random intercept); all variables are controlled for age and sex.

N = 1043; Model fit: CFI = .970; SRMR = .025; RMSEA = .073.

* p < .05. ** p < .01. *** p < .001.

In addition to the hypothesized paths, there was a significant positive path from aggressive behavior at T1 to peer problems at T2 (β = .07, 95% CI [.008, .136], p = .03) as well as from aggressive behavior at T2 to peer problems at T3 (β = .08, 95% CI [.001, .149], p = .048). The indirect paths from aggression T1 to aggression T3 through peer problems T2 (β = .01, 95% CI [.001, .029]) as well as from peer problems T1 to peer problems T3 through aggression T2 (β = .01, 95% CI [.001, .021]) were also significant. These findings attest to the mutually reinforcing negative effects of aggressive behavior and peer problems (for a meta-analytic review, see Reijntjes et al., 2011). Finally, a significant path from peer problems T2 to physical attractiveness T3 emerged (β = -.12, 95% CI [-.178, -.069], p < .001), indicating that – beyond temporal construct stability – problems in peer relationships are predictive for a negative development in facial attractiveness. In addition, physical attractiveness at T3 was indirectly predicted by aggressive behavior at T1 (β = -.01, 95% CI [-.019, -.001]), peer problems at T1 (β = -.06, 95% CI [-.091, -.034]), and physical attractiveness at T1 (β = .01, 95% CI [.0003, .0153]) through peer problems at T2.

To sum up, our main analysis supported our hypotheses: Physical attractiveness predicted subsequent peer problems, which in turn predicted aggressive behavior. Moreover, additional significant paths revealed reciprocal relations between peer problems and aggression over time, and peer problems as a potential predictor for lower physical attractiveness.

Separating Physical and Relational Aggression

In a further step, we broke down our measure of aggression into the two facets of physical and relational aggression to examine whether the associations would differ between the two forms.

For physical aggression the model showed a sufficient fit (χ^2[7] = 40.743, p < .001; CFI = .977; SRMR = .025; RMSEA = .068). Consistent with the main model, the hypothesized paths were significant (attractiveness T1 to peer problems T2: β = -.06, 95% CI [-.108, -.003], p = .04; peer problems T2 to physical aggression T3: β = .20, 95% CI [.121, .280], p < .001; corresponding indirect path: β = -.01, 95% CI [-.0137, -.0003]). However, we determined a moderation by sex, as indicated by a significant difference between the fully constrained and the fully unconstrained model ($\Delta\chi^2$[28] = 46.034, p = .02). We examined parameter differences to identify the paths that differed significantly in the two sex groups. Then, we computed a revised and final model in which only the paths that differed were allowed to vary. This revised model, presented in Figure III-3, showed a significantly better fit than the fully constrained

model ($\Delta\chi^2[7] = 28.453$, $p < .001$) and did not fit significantly worse than the fully unconstrained model ($\Delta\ \chi^2[21] = 14.687$, $p = .84$). Model fit for the revised model was good ($\chi^2[37] = 69.505$, $p < .001$; CFI = .970; SRMR = .041; RMSEA = .041). The path from attractiveness at T1 to peer problems at T2 did not reach statistical significance (girls: $\beta = -.06$, 95% CI [-.119, .000], $p = .051$, boys: $\beta = -.04$, 95% CI [-.087, .000], $p = .052$), probably caused by the reduced power due to halving of sample size in the multi-group analyses. The path from peer problems at T2 to physical aggression at T3 were significant for both sexes, but significantly stronger for boys ($\beta = .21$, 95% CI [.123, .297], $p < .001$) than for girls ($\beta = .15$, 95% CI [.050, .246], $p = .003$). In addition, there was a significant direct path from attractiveness at T1 to physical aggression at T3 in boys ($\beta = -.15$, 95% CI [-.235, -.068], $p < .001$), but not in girls ($\beta = -.05$, 95% CI [-.165, .071], $p = .44$). Finally, physical aggression predicted peer problems from T1 to T2 in both sexes (βs $\geq .06$, ps $< .05$), but from T2 to T3 only in girls ($\beta = .15$, 95% CI [.016, .291], $p = .03$; boys: $\beta = .04$, 95% CI [-.061, .137], $p = .45$).

For relational aggression, the model showed a sufficient fit ($\chi^2[7] = 33.466$, $p < .001$; CFI = .976; SRMR = .020; RMSEA = .060). The model is presented in Figure III-4. Consistent with the main model, the hypothesized paths were significant (attractiveness T1 to peer problems T2: $\beta = -.06$, 95% CI [-.111, -.005], $p = .03$; peer problems T2 to relational aggression T3: $\beta = .13$, 95% CI [.036, .214], $p = .006$; corresponding indirect path: $\beta = -.01$, 95% CI [-.0174, -.0004]). Although peer problems positively predicted subsequent relationally aggressive behavior (βs $\geq .12$, ps $< .05$), relational aggression did not predict peer problems (βs $\leq .06$, ps $\geq .10$). We did not find a significant moderation by sex ($\Delta\chi^2[28] = 35.628$, $p = .15$).

To sum up, our further analyses showed that the hypothesized paths held for the physical and relational form of aggressive behavior. However, significant sex differences emerged in our model for physical aggression. The final multi-group model revealed a negative direct path from attractiveness T1 to physical aggression T3 for boys. The proposed path from attractiveness T1 to peer problems T2 narrowly missed the significance threshold in both sex groups.

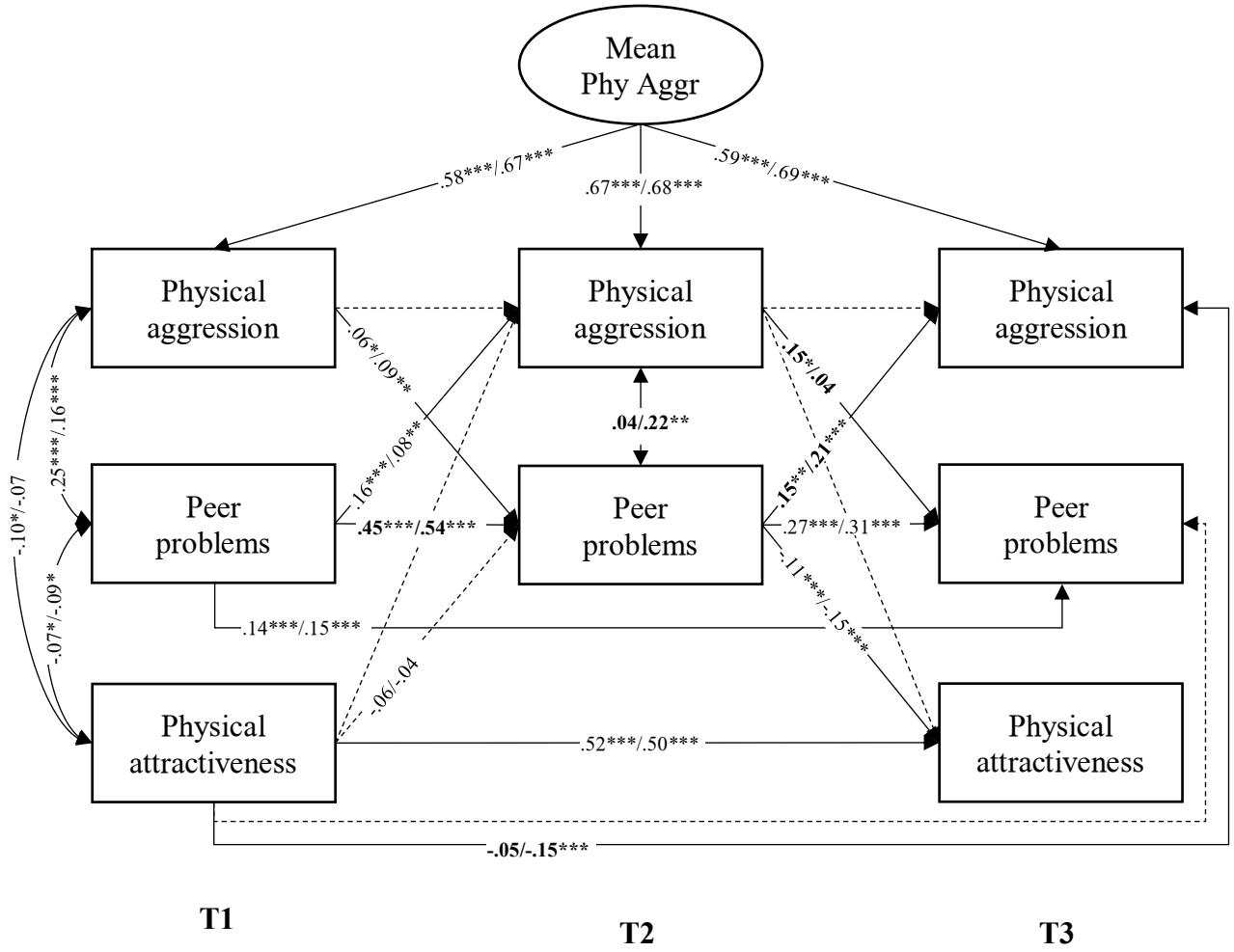

Figure III-3. Multi-group analysis for the specific form of physical aggression

Notes. Standardized coefficients are displayed; dashed lines display nonsignificant paths; mean physical aggression partials out between-person stability over time (random intercept); all variables are controlled for age; girls (n = 526) are presented on the left, boys (n = 517) on the right of the slash; significant sex differences are presented in bold.

N = 1043; Model fit: CFI = .970; SRMR = .041; RMSEA = .041.

* $p < .05$. ** $p < .01$. *** $p < .001$.

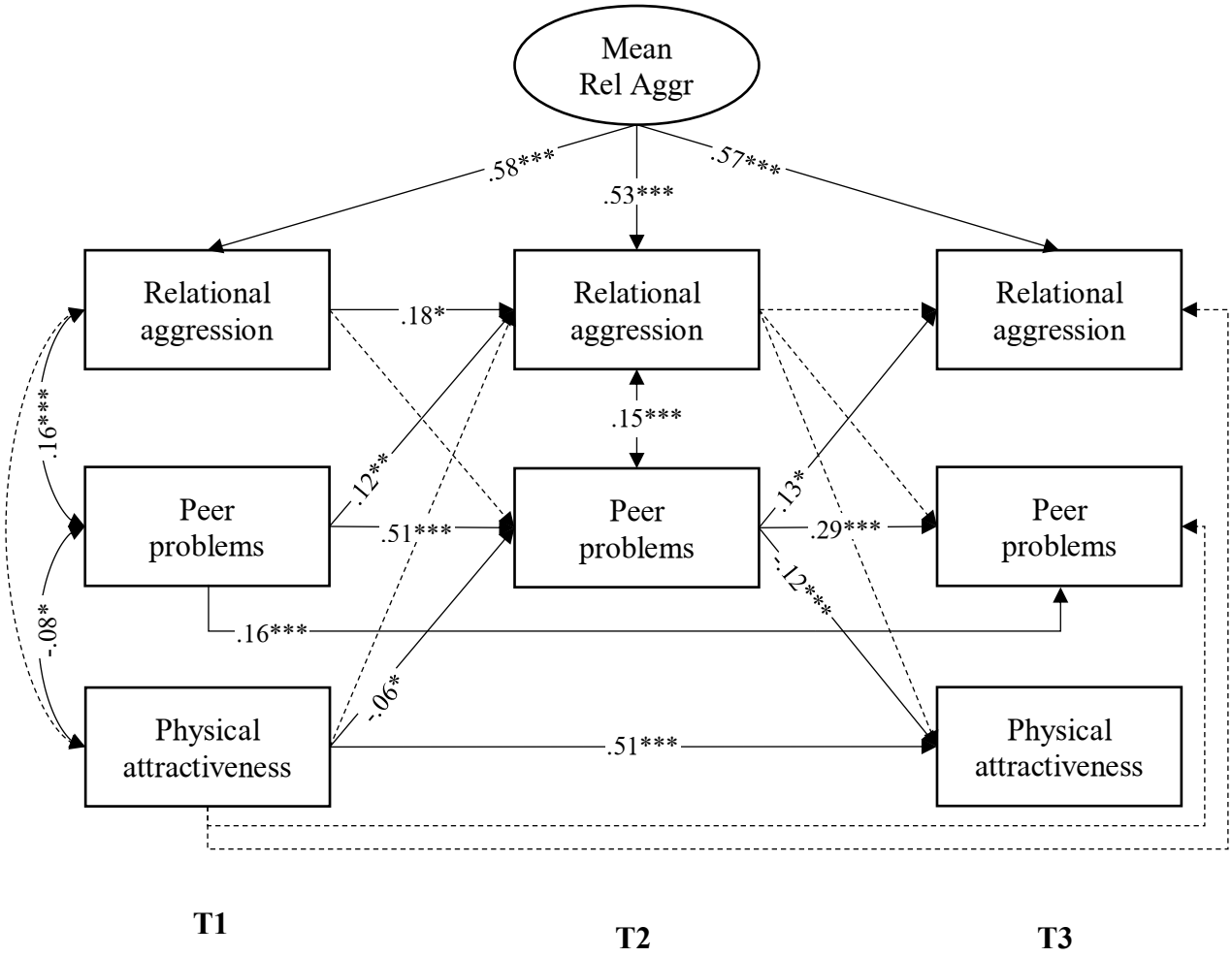

Figure III-4. Path analysis for the specific form of relational aggression

Notes. Standardized coefficients are displayed; dashed lines display nonsignificant paths; mean relational aggression partials out between-person stability over time (random intercept); all variables are controlled for age and sex.

N = 1043; Model fit: CFI = .976; SRMR = .020; RMSEA = .060.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Discussion

The aim of the present study was to examine the prospective relations of children's physical attractiveness, their experienced peer problems, and the development of their aggressive behavior over time. In a longitudinal study design covering three data waves in three years, we examined the following hypotheses: (1) Attractiveness is a negative predictor for subsequent peer problems, (2) peer problems predict subsequent aggressive behavior, and (3) low attractiveness indirectly predicts aggressive behavior through peer problems. We found evidence for all three hypotheses. Children who were rated less physically attractive by independent raters at T1 reported more problems in their peer relationships at T2, which in turn

predicted higher scores for aggressive behavior at T3. Thus, low physical attractiveness can be seen as a distal risk factor for the development of aggressive behavior in middle childhood through the social consequences attached to low attractiveness. These results applied to both girls and boys as we did not find a significant moderation by sex. There was no evidence for a positive association between attractiveness and aggression that is supposed by some authors (e.g., Borch et al., 2011).

It is noteworthy that our effect sizes were quite small. However, small effect sizes do not imply trivial effects (Cortina & Landis, 2009). Predicting the development of aggressive behavior is a complex issue, considering the many intra- and interpersonal factors that have been identified so far (for an overview, see Krahé, 2013). Although the bivariate correlation between attractiveness at T1 and aggression at T3 in our study was small (r = -.11), it was comparable in magnitude to other intrapersonal predictors of aggressive behavior, such as the hostile attribution bias (r = .13 for general population samples of nonreferred children and adolescents in a meta-analysis by de Castro, Veerman, Koops, Bosch, & Monshouwer, 2002). Thus, our study suggests that an individual's physical appearance should be considered alongside other intrapersonal characteristics that can have an impact on the development of aggressive behavior.

Our results are in line with socialization theories proposing that an individual's attractiveness elicits responses from the social environment that promote behaviors consistent with the attractiveness stereotype (e.g., Langlois et al., 2000). More attractive individuals are assumed to possess more positive traits ("What is beautiful is good"; Dion et al., 1972), whereas less attractive individuals are assumed to possess more negative characteristics ("What is ugly is bad"; e.g., Griffin & Langlois, 2006). These expectations lead to the differential treatment of individuals depending on their attractiveness. For example, previous research found that less attractive individuals face more problems in their peer relationships (e.g., Sweeting & West, 2010), a result we also found in our longitudinal study. The more unfavorable treatment of less attractive individuals by their social environment gives them more opportunities to develop in line with the stereotype, resulting in a self-fulfilling prophecy. In the case of aggressive behavior, a less attractive child may be expected to act more aggressively (e.g., Dion, 1973), so other children avoid contact with or reject that child, which in turn may trigger aggressive behavior in the child (e.g., Dodge et al., 2003; Rohlf et al., in press). Such self-fulfilling prophecy effects are especially problematic when accumulated across perceivers and time (for a review, see Madon, Willard, Guyll, & Scherr, 2011).

Further analyses revealed that the reported results held for both physical and relational aggression. However, for physical aggression significant sex differences emerged. For both sexes, experienced problems with peers at T2 were related to more physical aggression two years later at T3 (Hypothesis 2), but this relation was more pronounced in boys than in girls. The path from attractiveness at T1 to peer problems at T2 did not reach statistical significance for either sex (Hypothesis 1). This is most probably due to reduced power resulting from splitting the sample by sex. As a consequence, the proposed indirect effect from attractiveness to physical aggression via peer problems was also not significant in this analysis (Hypothesis 3). However, we found a negative direct effect from attractiveness at T1 to physical aggression T3 in boys, indicating that boys who were rated to be less attractive at T1 showed more physical aggression three years later at T3. As peer problems were not identified as a process underlying this path in the present data, only tentative suggestions can be offered to explain this link. One explanation could be that negative responses from adults rather than from peers may have played a role. There is evidence that less attractive children are also treated less favorably by their teachers (e.g., Dunkake, Kiechle, Klein, & Rosar, 2012) and even by their parents (e.g., Langlois, Ritter, Casey, & Sawin, 1995), and this effect seems somewhat stronger for boys. Furthermore, a recent study revealed that the short presentation of unattractive infant faces elicits facial muscle movement in adults that is associated with negative affect (Schein & Langlois, 2015). Thus, in future research it is important to not only examine the developmental impact of appearance-related treatment by peers, but also by significant adults.

Our path analyses included physical attractiveness not only at the first time point T1, but also at the final time point T3. This allowed us to examine the temporal development of physical attractiveness in our sample. Two major conclusions emerged from this analysis. First, physical attractiveness was found to be a relatively stable variable over the course of middle childhood. This result is in line with the few previous studies investigating the developmental stability of physical attractiveness during childhood (e.g., Yerkes & Pettijohn, 2008). Second, social experiences seem to affect the development of facial physical attractiveness, as indicated by a significant path from peer problems at T2 to attractiveness at T3. As children's attractiveness was assessed by raters unfamiliar with the children, we can rule out bias in attractiveness ratings based on the popularity of the child. Instead, one possible explanation may be the increased stress level resulting from peer problems (Peters, Riksen-Walraven, Cillessen, & de Weerth, 2011). Stress, in turn, is associated with a range of bodily changes, including skin problems (e.g., Kimyai-Asadi & Usman, 2001), overweight (e.g., Pervanidou & Chrousos, 2011), and physical illness (e.g., Cohen, Kessler, & Underwood Gordon, 1997).

Thus, ongoing peer problems may result in psychophysical stress that influences the physiological development of face and body towards lower attractiveness. For example, in a study by Rantala et al. (2013) the level of cortisol, a stress-related hormone, was negatively correlated with facial attractiveness in female adults. This possible pathway needs to be examined in future research including psychophysiological indicators of stress.

Strengths and Limitations

We believe our study has several strengths. First, we examined the effects of physical attractiveness in a longitudinal study design that controlled for past levels of the outcome variables, namely peer problems and aggressive behavior, and for initial correlations of attractiveness with these variables. This procedure goes some way toward a causal interpretation of the pathways (Marmor & Montemayor, 1977). Furthermore, we assessed physical attractiveness at the first and the final time point. This enabled us to examine not only the unidirectional effects attractiveness can have on subsequent social and behavioral development, but also potential bidirectional effects. Finally, we used different informants to assess the constructs of interest. This procedure reduces common method biases (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). Specifically, we used self-reports for measuring peer problems, teacher reports for aggressive behavior, in the form of both physical and relational aggression, and independent raters for assessing physical attractiveness.

At the same time, some limitations of our study have to be acknowledged. First, the photo material we used for the rating of physical attractiveness was extracted from video material collected for the purpose of behavioral observation (for further information, see Rohlf & Krahé, 2015). As a result, some deviations from perfect standardization emerged. For example, the children did not wear standardized clothes, and they did not show the exact same facial expressions. However, we increased the standardization as much as possible. We selected only photos in which the facial expressions were neutral, the children's eyes were visible, clothes were only minimally visible, and the photo quality was sufficient. Additionally, a black mask was added to increase the standardization between the photos and to reduce the noise within a photo (e.g., objects in the background were eliminated; see Figure 1). The photos we used and the resulting attractiveness ratings thus reflect the natural daily appearance of the children in the school context. Since our research was focused on children's peer problems within the school context, this feature may actually have enhanced the ecological validity of the photo material.

A second limitation is that we only used facial appearance as an indicator of physical attractiveness (i.e., facial attractiveness), ignoring other aspects of appearance that may also influence the perception of an individual's physical attractiveness, such as weight and body form (e.g., Connolly, Slaughter, & Mealey, 2004). In addition, attractiveness was measured only at T1 and T3, not at the intermediate time point. A full cross-lagged panel model would require all constructs to be assessed at all time points. Furthermore, attractiveness was assessed by adults, not by raters of the same age. Although previous research revealed that there is a consensus in attractiveness perception regardless of age (e.g., Slater et al., 1998), future research should establish whether peer rated facial attractiveness might have a greater predictive value for peer problems and aggression.

Finally, our results do not fully explain the link between physical attractiveness and the development of aggressive behavior. They suggest that peer problems play a critical role in this link, but it remains unclear why less attractive children experience more problems with their peers. Socialization theory suggests that low attractiveness leads to peer problems due to a more negative perception and more negative expectations by others about the unattractive child (e.g., Langlois et al., 2000). Because of the resulting unfavorable treatment, the individual develops in the expected negative way. However, it may be the case that less attractive individuals actually possess less positive characteristics, as assumed by evolutionary theories (e.g., Kanazawa & Kovar, 2004). In this sense, physical appearance would be a valid cue for gene quality ("good genes"; for a discussion, see Gangestad & Scheyd, 2005).

Implications

Despite these limitations, the present research indicates possible negative effects of appearance-based stereotypes on children's development of peer relationships and subsequent aggressive behavior. Future research is needed to decide whether the peer problems are a result of a genuine association of physical attractiveness and social skills, or a result of stereotype-based incorrect expectations that motivate others to avoid less attractive individuals. The latter assumption implies that interventions reducing the reliance on these stereotypes should lead to a decrease in the link between attractiveness and peer problems. As a consequence, the negative indirect effects of physical appearance on behavioral development would be diminished.

As children spend a lot of time using media for entertainment (Common Sense, 2015), one important starting point could be to change the way more versus less attractive individuals

are portrayed in the mass media. According to cultivation theory (for an overview, see Gerbner, Gross, Morgan, Signorielli, & Shanahan, 2002), media exposure shapes a users' ideology, values, and conception of the world. It is well known that media in Western cultures promote narrow beauty standards (e.g., Sypeck, Gray, & Ahrens, 2004). Even in media addressed at very young children, attractive characters are displayed as possessing more desirable traits (e.g., Bazzini, Curtin, Joslin, Regan, & Martz, 2010; Klein & Shiffman, 2006). Media could adopt a more active role in challenging the attractiveness stereotype. Another important approach may be to promote children's media literacy in challenging social stereotypes (Scharrer & Ramasubramanian, 2015).

Finally, our study suggests that social experiences can influence the development of physical attractiveness over time. To our knowledge there are no studies so far that examined this specific path. Future research is needed to replicate our findings and investigate possible mediating mechanisms, such as bodily changes caused by the psychophysical stress resulting from problematic peer interactions.

References

Asendorpf, J. B., & van Aken, M. A. G. (1993). Deutsche Versionen der Selbstkonzeptskalen von Harter [German version of the Harter's self-concept scales]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 25*, 64–86.

Baron, R. A., & Richardson, D. S. (1994). *Human aggression* (2nd ed.). New York, NY, US: Plenum Press.

Bazzini, D., Curtin, L., Joslin, S., Regan, S., & Martz, D. (2010). Do animated Disney characters portray and promote the Beauty-Goodness Stereotype? *Journal of Applied Social Psychology, 40*, 2687–2709. doi:10.1111/j.1559-1816.2010.00676.x

Benzeval, M., Green, M. J., & Macintyre, S. (2013). Does perceived physical attractiveness in adolescence predict better socioeconomic position in adulthood? Evidence from 20 years of follow up in a population cohort study. *PLOS ONE, 8*, e63975. doi:10.1371/journal.pone.0063975

Berkowitz, L., & Frodi, A. (1979). Reactions to a child's mistakes as affected by her/his looks and speech. *Social Psychology Quarterly, 42*, 420–425. doi:10.2307/3033814

Bobadilla, L., Metze, A. V., & Taylor, J. (2013). Physical attractiveness and its relation to unprovoked and reactive aggression. *Journal of Research in Personality, 47*, 70–77. doi:10.1016/j.jrp.2012.09.004

Borch, C., Hyde, A., & Cillessen, A. H. N. (2011). The role of attractiveness and aggression in high school popularity. *Social Psychology of Education, 14*, 23–39. doi:10.1007/s11218-010-9131-1

Cohen, S., Kessler, R. C., & Underwood Gordon, L. (1997). Strategies for measuring stress in studies of psychiatric and physical disorder. In S. Cohen, R. C. Kessler, & L. Underwood Gordon (Eds.), *Measuring stress. A guide for health and social scientists* (pp. 3–26). New York, NY, US: Oxford University Press.

Commisso, M., & Finkelstein, L. (2012). Physical attractiveness bias in employee termination. *Journal of Applied Social Psychology, 42*, 2968–2987. doi:10.1111/j.1559-1816.2012.00970.x

Common Sense. (2015). The common sense census: Media use by tweens and teens. Retrieved from https://www.commonsensemedia.org/sites/default/files/uploads/research/census_researchreport.pdf

Connolly, J. M., Slaughter, V., & Mealey, L. (2004). The development of preferences for specific body shapes. *Journal of Sex Research, 41*, 5–15. doi:10.1080/00224490409552209

Cortina, J. M., & Landis, R. S. (2009). When small effect sizes tell a big story, and when large effect sizes don't. In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends. Doctrine, verity and fable in the organizational and social sciences* (pp. 287–308). New York, NY, US: Routledge.

Crick, N. R. (1996). The role of overt aggression, relational aggression, and prosocial behavior in the prediction of children's future social adjustment. *Child Development, 67*, 2317–2327. doi:10.2307/1131625

de Castro, B. O., Veerman, J. W., Koops, W., Bosch, J. D., & Monshouwer, H. J. (2002). Hostile attribution of intent and aggressive behavior: A meta-analysis. *Child Development*, *73*, 916–934. doi:10.1111/1467-8624.00447

Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology, 24*, 285–290. doi:10.1037/h0033731

Dion, K. K. (1973). Young children's stereotyping of facial attractiveness. *Developmental Psychology, 9*, 183–188. doi:10.1037/h0035083

Dodge, K. A., Lansford, J. E., Burks, V. S., Bates, J. E., Pettit, G. S., Fontaine, R., & Price, J. M. (2003). Peer rejection and social information-processing factors in the development of aggressive behavior problems in children. *Child Development, 74*, 374–393. doi:10.1111/1467-8624.7402004

Dumas, J. E., Nilsen, W., & Lynch, A. M. (2001). How much does physical appearance say about the psychological adjustment of competent and dysfunctional children? *Journal of Clinical Child Psychology, 30*, 385–398. doi:10.1207/S15374424JCCP3003_10

Dunkake, I., Kiechle, T., Klein, M., & Rosar, U. (2012). Schöne schüler, schöne Noten? Eine empirische Untersuchung zum Einfluss der physischen Attraktivität von Schülern auf die Notenvergabe durch das Lehrpersonal [Good looks, good grades? An empirical analysis of the influence of students' physical attractiveness on grading by teachers]. *Zeitschrift für Soziologie, 41*, 142–161.

Eagly, A. H., Ashmore, R. D., Makhijani, M. G., & Longo, L. C. (1991). What is beautiful is good, but…: A meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin, 110*, 109–128. doi:10.1037/0033-2909.110.1.109

Enders, C., & Bandalos, D. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling, 8*, 430–457. doi:10.1207/S15328007SEM0803_5

Gadermann, A. M., Guhn, M., & Zumbo, B. D. (2012). Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. *Practical Assessment, Research & Evaluation*, *17*(3), 1–13.

Gangestad, S. W., & Scheyd, G. J. (2005). The evolution of human physical attractiveness. *Annual Review of Anthropology, 34*, 523–548. doi:10.1146/annurev.anthro.33.070203.143733

Gerbner, G., Gross, L., Morgan, M., Signorielli, N., & Shanahan, J. (2002). Growing up with television: Cultivation processes. In J. Bryant & D. Zillmann (Eds.), *Media effects. Advances in theory and research* (2nd ed., pp. 43–67). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Goswami, H. (2012). Social relationships and children's subjective well-being. *Social Indicators Research, 107*, 575–588. doi:10.1007/s11205-011-9864-z

Griffin, A. M., & Langlois, J. H. (2006). Stereotype directionality and attractiveness stereotyping: Is beauty good or is ugly bad? *Social Cognition, 24*, 187–206. doi:10.1521/soco.2006.24.2.187

Gueguen, N., Lourel, M., Charron, C., Fischer-Lokou, J., & Lamy, L. (2009). A web replication of Snyder, Decker, and Bersheid's (1977) experiment on the self-fulfilling nature of social stereotypes. *The Journal of Social Psychology, 149*, 600–602. doi:10.1080/00224540903238503

Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. P. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, *20*, 102–116. doi:10.1037/a0038889

Hayes, A. F., & Scharkow, M. (2013). The relative trustworthiness of inferential tests of the indirect effect in statistical mediation analysis: Does method really matter? *Psychological Science*, *24*, 1918–1927. doi:10.1177/0956797613480187

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

Kanazawa, S., & Kovar, J. L. (2004). Why beautiful people are more intelligent. *Intelligence, 32*, 227–243. doi:10.1016/j.intell.2004.03.003

Kimyai-Asadi, A., & Usman, A. (2001). The role of psychological stress in skin disease. *Journal of Cutaneous Medicine and Surgery*, *5*, 140–145. doi:10.1007/s102270000016

Klein, H., & Shiffman, K. S. (2006). Messages about physical attractiveness in animated cartoons. *Body Image, 3*, 353–363. doi:10.1016/j.bodyim.2006.08.001

Krahé, B. (2013). *The social psychology of aggression* (2nd ed.). Hove, UK: Psychology Press.

Langlois, J. H., & Downs, A. C. (1979). Peer relations as a function of physical attractiveness: The eye of the beholder or behavioral reality? *Child Development, 50*, 409–418. doi:10.1111/j.1467-8624.1979.tb04122.x

Langlois, J. H., Kalakanis, L., Rubenstein, A. J., Larson, A., Hallam, M., & Smoot, M. (2000). Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychological Bulletin, 126*, 390–423. doi:10.1037//0033-2909.126.3.390

Langlois, J. H., Ritter, J. M., Casey, R. J., & Sawin, D. B. (1995). Infant attractiveness predicts maternal behaviors and attitudes. *Developmental Psychology, 31*, 464–472. doi:10.1037/0012-1649.31.3.464

Leary, M. R., Twenge, J. M., & Quinlivan, E. (2006). Interpersonal rejection as a determinant of anger and aggression. *Personality and Social Psychology Review, 10*, 111–132. doi:10.1207/s15327957pspr1002_2

Ma, F., Xu, F., & Luo, X. (2015). Children's and adults' judgments of facial trustworthiness: The relationship to facial attractiveness. *Perceptual and Motor Skills*, *121*, 179–198. doi:10.2466/27.22.PMS.121c10x1

Madon, S., Willard, J., Guyll, M., & Scherr, K. C. (2011). Self-Fulfilling Prophecies: Mechanisms, power, and links to social problems. *Social and Personality Psychology Compass, 5*, 578–590. doi:10.1111/j.1751-9004.2011.00375.x

Marmor, G. S., & Montemayor, R. (1977). The cross-lagged panel design: A review. *Perceptual and Motor Skills, 45*, 883–893. doi:10.2466/pms.1977.45.3.883

Mazzella, R., & Feingold, A. (1994). The effects of physical attractiveness, race, socioeconomic status, and gender of defendants and victims on judgments of mock jurors: A meta-analysis. *Journal of Applied Social Psychology, 24*, 1315–1338. doi:10.1111/j.1559-1816.1994.tb01552.x

Muthén, L. K., & Muthén, B. O. (1998-2015). *Mplus user's guide* (7th ed.). Los Angeles, CA, US: Muthén & Muthén.

NGWIN. (2016). PicPick [Computer Software]. Retrieved from http://ngwin.com/picpick

Patzer, G. L. (2006). *The power and paradox of physical attractiveness*. Boca Raton, FL, US: Brown Walker Press.

Pervanidou, P., & Chrousos, G. P. (2011). Stress and obesity/metabolic syndrome in childhood and adolescence. *International Journal of Pediatric Obesity*, *6*, 21–28. doi:10.3109/17477166.2011.615996

Peters, E., Riksen-Walraven, J. M., Cillessen, Antonius H. N., & de Weerth, C. (2011). Peer rejection and HPA activity in middle childhood: Friendship makes a difference. *Child Development*, *82*, 1906–1920. doi:10.1111/j.146

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *The Journal of Applied Psychology, 88*, 879–903. doi:10.1037/0021-9010.88.5.879

Rantala, M. J., Coetzee, V., Moore, F. R., Skrinda, I., Kecko, S., Krama, T.,. . . Krams, I. (2013). Facial attractiveness is related to women's cortisol and body fat, but not with immune responsiveness. *Biology Letters*, *9*, 20130255. doi:10.1098/rsbl.2013.0255

Rauer, W., & Schuck, K. D. (2003). *FEESS 3-4: Fragebogen zur Erfassung emotionaler und sozialer Schulerfahrungen von Grundschulkindern dritter und vierter Klassen* [FEESS 3-4: Questionnaire on social and emotional experiences at school of elementary school children in third and fourth grade]. Göttingen, D: Beltz Test GmbH.

Rauer, W., & Schuck, K. D. (2004). *FEESS 1-2: Fragebogen zur Erfassung emotionaler und sozialer Schulerfahrung von Grundschulkindern erster und zweiter Klassen* [FEESS 1-2: Questionnaire on social and emotional experiences at school of elementary school children in first and second grade]. Göttingen, D: Beltz Test GmbH.

Reijntjes, A., Kamphuis, J. H., Prinzie, P., Boelen, P. A., van der Schoot, M., & Telch, M. J. (2011). Prospective linkages between peer victimization and externalizing problems in children: A meta-analysis. *Aggressive Behavior, 37*, 215–222. doi:10.1002/ab.20374

Rohlf, H., Busching, R., & Krahé, B. (in press). Longitudinal links between maladaptive anger regulation, peer problems, and aggression in middle childhood. *Merrill-Palmer Quarterly*.

Rohlf, H., & Krahé, B. (2015). Assessing anger regulation in middle childhood: Development and validation of a behavioral observation measure. *Frontiers in Psychology*, *6*, 453. doi:10.3389/fpsyg.2015.00453

Rosen, L. H., & Underwood, M. K. (2010). Facial attractiveness as a moderator of the association between social and physical aggression and popularity in adolescents. *Journal of School Psychology, 48*, 313–333. doi:10.1016/j.jsp.2010.03.001

Rosen, L. H., Underwood, M. K., & Beron, K. J. (2011). Peer victimization as a mediator of the relation between facial attractiveness and internalizing problems. *Merrill-Palmer Quarterly, 57*, 319–347. doi:10.1353/mpq.2011.0016

Scharrer, E., & Ramasubramanian, S. (2015). Intervening in the media's influence on stereotypes of race and ethnicity: The role of media literacy education. *Journal of Social Issues, 71*, 171–185. doi:10.1111/josi.12103

Schein, S. S., & Langlois, J. H. (2015). Unattractive infant faces elicit negative affect from adults. *Infant Behavior & Development, 38*, 130–134. doi:10.1016/j.infbeh.2014.12.009

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online, 8*, 23–74.

Serketich, W. J., & Dumas, J. E. (1997). Adults' perceptions of the behavior of competent and dysfunctional children based on the children's physical appearance. *Behavior Modification, 21*, 457–469. doi:10.1177/01454455970214004

Slater, A., von der Schulenburg, C., Brown, E., Badenoch, M., Butterworth, G., Parsons, S., & Samuels, C. (1998). Newborn infants prefer attractive faces. *Infant Behavior and Development*, *21*, 345–354. doi:10.1016/S0163-6383(98)90011-X

Smith, G. J. (1985). Facial and full-length ratings of attractiveness related to the social interactions of young children. *Sex Roles, 12*, 287–293. doi:10.1007/BF00287595

Sweeting, H., & West, P. (2010). Being different: Correlates of the experience of teasing and bullying at age 11. *Research Papers in Education, 16*, 225–246. doi:10.1080/02671520110058679

Sypeck, M. F., Gray, J. J., & Ahrens, A. H. (2004). No longer just a pretty face: Fashion magazines' depictions of ideal female beauty from 1959 to 1999. *The International Journal of Eating Disorders, 36*, 342–347. doi:10.1002/eat.20039

Umberson, D., Chen, M. D., House, J. S., Hopkins, K., & Slaten, E. (1996). The effect of social relationships on psychological well-being: Are men and women really so different? *American Sociological Review, 61*, 837–857. doi:10.2307/2096456

Umberson, D., & Montez, J. K. (2010). Social relationships and health: A flashpoint for health policy. *Journal of Health and Social Behavior, 51*, 54-66. doi:10.1177/0022146510383501

Yerkes, M. J., & Pettijohn, T. F. (2008). Developmental stability of perceived physical attractiveness from infancy to young adulthood. *Social Behavior and Personality*, *36*, 691–692.

Zakin, D. F. (1983). Physical attractiveness, sociability, athletic ability, and children's preference for their peers. *The Journal of Psychology, 115*, 117–122. doi:10.1080/00223980.1983.9923606

Zimmer-Gembeck, M. J., Geiger, T. C., & Crick, N. R. (2005). Relational and physical aggression, prosocial behavior, and peer relations: Gender moderation and bidirectional associations. *The Journal of Early Adolescence, 25*, 421–452. doi:10.1177/0272431605279841

6

General Discussion

The primary aim of this doctoral dissertation was to identify intrapersonal risk factors of aggressive behavior in child development. The prominent SIP model by Crick and Dodge (1994) was used as a framework for examining potential risk factors in the cognitive, affective, and appearance-related domain. Individual differences in these domains are assumed to influence SIP and, thus, may be important for understanding the occurrence and the development of aggressive behavior. Each empirical study presented in this dissertation aimed to investigate the predictive impact of one domain on childhood aggression, using longitudinal data from the PIER-Project. Specifically, Paper 1 examined *anger regulation* as an affective factor, Paper 2 examined *theory of mind* as a cognitive factor, and Paper 3 examined *physical attractiveness* as an appearance-related factor of aggressive behavior. An additional major goal of this work was to develop and validate a behavioral observation measure of anger regulation for the use in longitudinal studies. Thus, the following general discussion, firstly, discusses the development of this measure, secondly, summarizes the results of the papers in terms of the identification of intrapersonal risk factors of aggression, and thirdly, discusses general strengths as well as limitations of the studies. The discussion ends with conceptual and practical implications that can be inferred from the presented findings.

6.1 Assessing Anger Regulation in Childhood

Assessing children's emotion regulation in general and their anger regulation in particular is mostly done by using self-report measures (Aldao et al., 2010). Although such measures have notable advantages, their use in childhood is fraught with problems (Underwood, 1997). The main concern is that children's report about their use of regulation strategies in hypothetical situations may considerably differ from their use in actual emotion-eliciting situations (e.g., Parker et al., 2001). Rohlf and Krahé (2015) developed a behavioral observation assessment of anger regulation strategy use in an actual anger-eliciting situation to overcome these problems. They examined the same children who took part in the studies reported in this doctoral dissertation, but only cross-sectionally at T1. The children were given a dexterity task that was virtually impossible to complete (stacking ten wooden blocks to form a stable tower with two blocks rounded on one side). This procedure was able to elicit anger, an essential prerequisite for the investigation of anger regulation. Anger regulation was assessed by coding five specific regulation strategies, namely solution orientation (adaptive strategy) as well as visual and verbal focus on the frustrating stimuli, resignation, and venting the anger (maladaptive strategies). This measure was found to be reliable as indicated by high interrater agreement, and valid as

indicated by concurrent correlations with the conceptually similar construct anger reactivity (construct validity), and aggression and social rejection (criterion validity). In a follow-up study, the measure was also found to have predictive validity (Kirsch et al., 2015).

Since the present doctoral dissertation sought to investigate intrapersonal risk factors for the development of aggressive behavior, it was necessary to repeat the anger regulation assessment to control for potential reciprocal effects over time (e.g., Selig & Little, 2012). However, using the exact same anger-eliciting task again is inherently problematic as possible memory and training effects could increase task reactivity, which in turn threatens the internal validity of the measure (Schaie & Hofer, 2001). Furthermore, the children got older and, thus, the original task with wooden blocks may not be age-appropriate anymore. Therefore, the major goal of Paper 1 was to further develop the assessment by Rohlf and Krahé (2015) in a way that the anger-eliciting task is sufficiently different from – but conceptually equivalent to – the original task. Thus, at T3, the children were given a very similar dexterity task (stacking seven dice to form a tower with two tetrahedrons that were virtually impossible to stack on top of each other). In addition, the T1 coding scheme was used for the behavioral observation at T3 with only slight deviations (see Paper 1), resulting in the same five coded anger regulation strategies. This procedure was found to be reliable as indicated by high interrater agreement, and the new task was effective in eliciting angry feelings.

The first aim was to establish at least metric measurement invariance between the two measures to ensure conceptual equivalence. In line with this aim, the results revealed partial metric invariance to be the highest level of longitudinal measurement invariance for the two assessments (factor loadings of the strategy solution orientation were freed). This level of invariance allows to compare regression slopes in longitudinal studies (Chen, 2007). However, it does not allow to compare latent mean differences across time as were enabled by establishing scalar invariance (Widaman & Reise, 1997). Thus, the new measure cannot be viewed as a strict parallel version of the original measure, but as sufficiently equivalent to ensure that the measured constructs have the same unit of measurement.

As a second aim, a substantial correlation between the two measures was expected since emotion regulation is considered to be an individual difference variable that is relatively stable over time (e.g., Cole et al., 1994). Indeed, the results revealed a correlation of moderate size, indicating a moderate construct stability over the course of three years from T1 to T3. In addition, as the T1 task was found to be construct valid (Rohlf & Krahé, 2015) and emotion

regulation is conceptualized as a stable variable (Cole et al., 1994), this result may also be interpreted as an indication that the T3 measure is construct valid (Cronbach & Meehl, 1955).

To validate the newly developed assessment at T3 correlations with variables, which are supposed to be associated with emotion regulation, were examined. Specifically, the correlation with aggressive behavior (teacher-report, self-report, and in-situ behavior), peer problems (teacher-, self-, and parent-report), and conduct problems (parent-report) were tested. In addition, the original T1 assessment was included in these analyses. Both the T1 and the T3 measures were found to be significantly correlated with the validation constructs cross-sectionally (concurrent criterion validity). Moreover, the T1 measure was prospectively correlated with all of the validation constructs at T3 (predictive criterion validity), with an exception of teacher-reported aggression. Thus, the findings support the validity of the T3 measure (concurrent) and provide additional support for the validity of the T1 measure (concurrent and predictive).

In sum, Paper 1 was able to present a newly developed, reliable, and valid assessment of children's anger regulation strategies that is sufficiently equivalent to the original assessment developed by Rohlf and Krahé (2015). This work adds to the methodological repertoire of assessing anger regulation in children in the age of 6 to 13 years, and provides the opportunity to study the use of actual anger regulation strategies longitudinally.

In addition, the study emphasized several important points about the nature of behavioral observation measures for assessing emotion regulation. For example, behavioral observation is by definition only applicable for regulation strategies that are observable. Thus, such measures are not able to capture the use of mental strategies, such as cognitive reappraisal. Furthermore, the use of arranged situations for eliciting specific emotions usually constrains children's possibilities of action. For example, in the presented study, children were not given the opportunity to select or modify the situation. Thus, researchers should be aware that observed regulation behavior does not exhaustively represent the subject's regulatory competencies.

Another aspect that this study highlighted is the difficulty to induce an isolated emotion in an arranged setting. In the present study, the children were asked after the anger-eliciting situation to what extent they experienced anger and sadness. They reported significantly more anger than sadness. However, this does not imply that sadness was absent during the task. In fact, descriptive statistics showed that some children experienced at least a little sadness (M = 1.84 at T1 and M = 1.65 at T3, respectively, on a scale ranging from 1 to 3). This finding,

however, is not a weakness of the specific task because emotional experiences in general are strongly interrelated and can take place simultaneously (e.g., Trampe, Quoidbach, & Taquet, 2015). Thus, when trying to elicit a specific emotion, one must be aware that other emotions can be triggered, too. Consequently, it may be that not all observed and coded behaviors are actually behaviors that regulate the intended emotion. In terms of the present study, for some children, the observed strategies may regulate not only anger, but also other emotions. For example, the children may experience the test situation as inherently anxiety-inducing as the situation was similar to performance-oriented tests. It may also be that the experienced and anticipated failure elicited sadness. On the other hand, some children may experience joy during the dexterity task. In conclusion, the induction of a specific, isolated emotion is generally difficult to manage. Thus, researchers should be aware that behavioral observation assessments for the regulation of a single emotion may not be limited to the intended emotion. To address this problem, future studies could engage in asking for emotional experiences of diverse emotions when using such behavioral observation assessments.

6.2 Risk Factors of Aggressive Behavior in Childhood

6.2.1 Maladaptive Anger Regulation

Based on the developed and validated assessments described above, anger regulation was selected to represent an affective developmental factor of children's aggressive behavior. Although the emotion of anger is closely linked to aggression (e.g., Stemmler, 2010), the majority of past research examined emotion regulation in general rather than anger regulation in particular. According to the SIP model (Crick & Dodge, 1994) and its extended version by Lemerise and Arsenio (2000), emotional processes can shape social behavior by affecting SIP. For example, anger may color the interpretation of social cues (SIP step 2), resulting in higher probabilities to attribute hostile intentions to others' ambiguous behavior, which in turn can lead to aggressive behavior (de Castro et al., 2002). Thus, a maladaptive regulation of anger would result in persistent or even increased levels of anger (e.g., Denson et al., 2012), leading to biased SIP and, in turn, to an increasing probability to show aggression.

In line with this reasoning, it was hypothesized that maladaptive anger regulation is predictive of subsequent aggressive behavior (i.e., a prospective risk factor). In Paper 1, maladaptive anger regulation was clearly correlated with diverse measures of children's aggressive behavior concurrently (teacher-report, self-report, and in-situ behavior). On a predictive level, maladaptive anger regulation at T1 was only associated with the self-report

and in-situ behavior three years later at T3, but not with the teacher-report. More elaborated analyses were conducted to test whether maladaptive anger regulation is predictive of aggression when controlling for stability and initial correlations of the two constructs as well as potential reciprocal effects (i.e., a cross-lagged panel model, see Appendix A). These analyses were done using the teacher-report of aggression because only this measure was used at both measurement points. In addition, two models were considered: the first for overall aggression, and the second to differentiate between physical and relational aggression. The results revealed that, even though the constructs were correlated with one another concurrently, maladaptive anger regulation was not able to predict subsequent aggression, neither the overall score nor one of the two forms. Thus, in this research, there is no evidence that a maladaptive regulation of anger is actually a prospective risk factor of aggressive behavior in children. These results are in contrast to previous longitudinal research finding that emotion regulation is predictive of aggression, even when controlled for past levels of the outcome (for a review, see Röll et al., 2012).

One reason why maladaptive anger regulation was not found to be predictive of subsequent aggression in this study may be that the anger regulation measure was constrained in the observation of relevant regulation strategies, as described above. In particular, children's use of mental strategies, such as cognitive reappraisal, was not able to be captured in the behavioral observation. However, such mental strategies become considerably important when children get older (e.g., McRae et al., 2012; see Chapter 2.2.1). It may be that internal anger regulation is more important for the prediction of aggressive behavior of school-aged children than their observable external strategies.

Another probable reason may be the context in which the anger-eliciting task was embedded. Aggressive behavior is a social behavior, that is, a behavior that involves other people. The anger-eliciting task used for the behavioral observation, however, was more of a non-social, but performance situation. It may be that children who were found to be maladaptive regulators in the arranged performance-like situation are not necessarily maladaptive regulators in peer-related situations. This notion is consistent with recent research revealing that selection and perceived effectiveness of emotion regulation strategies can vary depending on social context (e.g., English, Lee, John, & Gross, 2017; Waters & Thompson, 2016) and personal motives (for a review, see Tamir, 2016). That is, for example, children may be more motivated to down regulate their anger expression in a social compared to a non-social situation. Thus, since peer relationships gain in importance with age (Lam et al., 2014), anger regulation in peer

situations may be more meaningful for the prediction of aggression of school-aged children than their anger regulation in non-social situations.

6.2.2 Deficits in Theory of Mind

As a cognitive developmental factor of childhood aggression, theory of mind was examined; that is, the ability to attribute mental states to the self or to others. Similar to maladaptive anger regulation, deficits in theory of mind are associated with biased SIP. For example, Choe et al. (2013) found that preschoolers with inferior theory of mind tend to attribute hostile intentions to others. As a result of such problems in SIP, it can be assumed that the probability to show aggressive behavior is increased (e.g., de Castro et al., 2002). However, some researchers argue that superior theory of mind may actually be beneficial for those forms of aggressive behavior that require the ability to manipulate others effectively (i.e., relational aggression; Renouf et al., 2010; Sutton et al., 1999b). Thus, the link between theory of mind skills and the development of aggression may differ for physical versus relational aggression.

In line with this reasoning, it was hypothesized that theory of mind is predictive of subsequent aggression. Specifically, it was assumed that this link is negative for physical aggression, whereas for relational aggression, both positive and negative associations could be expected depending on the theoretical point of view. The longitudinal analyses reported in Paper 2 support the hypotheses. Theory of mind was predictive of both physical and relational aggression, from T1 to T2 and from T2 to T3, and the direction of effects was negative. That is, the study found evidence for inferior theory of mind to function as a risk factor of aggressive behavior in childhood. This result is consistent with the social skills deficit view supposed by the SIP model (Crick & Dodge, 1994) and previous research (e.g., O'Toole et al., 2017; Kokkinos et al., 2016). At the same time, there was no evidence that superior theory of mind could foster relational forms of aggression as proposed and found by some authors (e.g., Sutton et al., 1999a, 1999b).

There may be some important moderators in the link between theory of mind and aggressive behavior that may explain these divergent findings. For example, Renouf et al. (2010) found a positive association between theory of mind and relational aggression, but only in children who scored low or average in prosocial behavior. For children with high levels of prosocial behavior, the authors found no correlation. Thus, individual differences in specific variables (e.g., prosocial behavior) may moderate the effects presented in Paper 2, and should be considered in future studies.

Additionally, future studies should focus on potential mediating variables in the link between theory of mind and aggression. According to the SIP model (Crick & Dodge, 1994), biases in SIP (e.g., hostile attribution bias) could function as such proximal mediators. However, other mediating mechanisms also are possible. For example, as a study with Greek elementary school students by Kokkinos et al. (2016) suggests, moral disengagement could play an important role. Moral disengagement describes cognitive processes that disengage moral and ethical standards from one's own behavior. They found that deficits in theory of mind indirectly predict increased relational aggression through increased moral disengagement, especially in boys. That is, children who have problems with theory of mind have a higher probability to morally disengage from their behavior, and that, in turn, may foster aggressive acts.

In addition to the hypotheses, examining deficits in theory of mind as a risk factor of aggressive behavior, it was assumed that involvement in aggression can predict subsequent theory of mind. That is, reciprocal effects over time were hypothesized between aggression and theory of mind. This reasoning was based on the conceptualization of the SIP model by Crick and Dodge (1994) emphasizing social experiences to be able to shape individual differences in SIP. The involvement in frequent aggressive acts can be considered to be such a social experience. At the same time, theory of mind is considered to be a core ability that is involved in most of the SIP steps. Thus, the engagement in aggressive behavior may be predictive of theory of mind development. The results presented in Paper 2 are partially consistent with that notion. Relationally, but not physically aggressive behavior at T1 was predictive of theory of mind at T2. Children who were more relationally aggressive showed inferior theory of mind skills nine months later than children, who were less relationally aggressive. However, physical aggression at T1 did not predict theory of mind at T2. During school age, peer relationships gain in importance (Lam et al., 2014). Relational aggression specifically aims to damage social relationships and, thus, the involvement in relationally aggressive behavior may trigger social experiences that are more influential for socio-cognitive development in this age.

The integration of the findings (i.e., the reciprocal effects between aggressive behavior and theory of mind over time) suggests potential self-reinforcing cycles of aggression (e.g., Jung, Krahé, Bondü, Esser, & Wyschkon, 2016). That is, early aggression may foster later aggression through the intermediate disadvantageous development of socio-cognitive skills. In Appendix B, this notion was tested by computing indirect paths (1) from relational aggression T1 to relational aggression T3 via theory of mind T2, and (2) from relational aggression T1 to physical aggression T3 via theory of mind T2. Both tested indirect paths were of small size, but statistically significant. Thus, relationally aggressive children were at a higher risk to develop

less skilled in their theory of mind, which in turn triggered subsequent aggressive behavior (physically and relationally). This finding highlights the fact that the examined intrapersonal risk factors can be both predictor and outcome of aggressive behavior.

6.2.3 Lack of Physical Attractiveness

Physical attractiveness was examined as a distal appearance-related developmental factor of childhood aggression. It was assumed that low attractiveness would lead to increased aggression caused by problematic peer relationships. As the perception of attractiveness is linked to the generation of stereotypical expectations of an individual, that individual should experience corresponding treatment by his or her social environment, reflected in the degree of experienced peer problems. Specifically, attractive children are assumed to trigger mainly positive expectations ("What is beautiful is good"; e.g., Dion et al., 1972), whereas unattractive children are assumed to trigger negative expectations more likely ("What is ugly is bad"; e.g., Griffin & Langlois, 2006). As a consequence, less attractive children are expected to receive a less favorable treatment by their social environment than their more attractive counterparts. These appearance-dependent social experiences are supposed to be stored in the memory (i.e., in the database of the SIP model, Crick & Dodge, 1994), and can be retrieved in virtually all SIP steps. For example, children who are used to experience rejection by their peers may interpret others' behavior in more hostile ways (SIP step 2), or hold worse outcome expectations, resulting in decisions for actually less beneficial social responses in a given situation (SIP step 5).

In line with this reasoning, it was hypothesized that (1) physical attractiveness is a negative predictor of subsequent peer problems, and (2) these peer problems are predictive of subsequent aggression, resulting in (3) a negative indirect path from attractiveness to aggression via peer problems. The study reported in Paper 3 was designed to test these assumptions. The results supported the hypotheses: Physical attractiveness at T1 was a negative predictor of peer problems at T2, which in turn was a positive predictor of aggressive behavior at T3. The corresponding indirect path was also statistically significant. These results are in line with previous research, demonstrating a negative correlation between children's physical attractiveness and their involvement in aggressive behavior (e.g., Langlois & Downs, 1979; Serketich & Dumas, 1997) and problematic peer relationships (e.g., Sweeting & West, 2010). At the same time, the results are not consistent with the notion of some authors that more

attractive compared to less attractive individuals show a greater tendency for aggressive behavior as their attractiveness protects them from social sanctions (Borch et al., 2011).

In conclusion, the findings suggest that problematic peer relationships, which result from low attractiveness, are critical for the link between a less appealing appearance and the development of social behavior. However, it remains unclear why less attractive children experience more peer problems than more attractive children. As discussed above, this phenomenon may result from the mere activation of the attractiveness stereotype. Unattractiveness compared to attractiveness triggers the social environment to generate less positive expectations about the personality and behavior of an individual (for a meta-analytic review, see Langlois et al., 2000). In this sense, unattractive individuals experience unfavorable social interactions solely caused by their appearance. On the other hand, a genuine correlation between physical attractiveness and personality could exist that may explain the peer problems of unattractive children. For example, less attractive children compared to their more attractive counterparts may genuinely be less social competent as suggested by some evolutionary theories (e.g., Kanazawa & Kovar, 2004). Moreover, a combination of both theoretical explanations is possible: Less attractive individuals may indeed possess more negative qualities on a genetic basis, and the negative expectations held by others about that individual foster the further development of these negative qualities. This possibility is in line with the recent nature versus nurture debate, which suggests that nature (i.e., genetics) and nurture (i.e., environment) influence one another (Eagly & Wood, 2013).

Another mechanism that may play a role in this context is stereotype threat (for a review, see Inzlicht & Schmader, 2012). Since physical attractiveness is a very salient cue (Olson & Marshuetz, 2005) and associated with a specific stereotype (Dion et al., 1972), children may be aware of their own level of attractiveness and the corresponding expectations others hold about them. That, in turn, could guide their behavior. For example, it may be that less attractive children suppose that their peers expect them to act socially inadequate. These children may be afraid to confirm this stereotype by their behavior; that is, they feel stereotype threat. In their proposed process model, Schmader, Johns, and Forbes (2008) argue that stereotype threat produces physiological stress responses, activates monitoring processes, and leads to efforts suppressing negative cognitions and emotions. These resource-intensive processes undermine actual performance (here, social adequate behavior). Recent studies suggest that stereotype threat is already observable in elementary school students (e.g., Hermann & Vollmeyer, 2016). However, the research on stereotype threat is predominantly focused on sex, race, and age

stereotypes, the attractiveness stereotype has not been sufficiently addressed yet. Thus, future studies are needed to test this notion.

6.2.4 Further Considerations

In sum, the presented studies suggest that deficits in theory of mind and a lack of physical attractiveness can function as risk factors for the development of aggressive behavior in childhood. However, although correlated with aggressive behavior, maladaptive anger regulation was not found to be a prospective risk factor of aggression. These findings suggest that various intrapersonal domains contribute to the prediction of aggressive behavior and should be considered when it comes to the conceptualization of prevention programs (see Chapter 6.4). Furthermore, even though it was not explicitly tested in the current studies, all intrapersonal risk factors were supposed to have one aspect in common, namely, the potential impact on SIP that was assumed to be critical in the link between the risk factors and the occurrence of aggression (Crick & Dodge, 1994).

The examined variables anger regulation, theory of mind, and physical attractiveness are not meant to exhaustively represent the three intrapersonal domains of affect, cognition, and appearance. Considering the SIP model by Crick and Dodge (1994), there are many intrapersonal characteristics that may be able to color SIP, directly and indirectly through the database, and, thus, may be correlated with the occurrence and development of aggressive behavior. In their extension of the SIP model, Lemerise and Arsenio (2000) proposed that – besides emotion regulation – temperament, emotionality, and mood states may be relevant in the affective domain. For example, mood can affect the encoding and interpretation of social cues (SIP steps 1 and 2) as mood-congruent information usually receive more attention. In the cognitive domain, the general tendency to attribute hostile intentions to others is one of the most frequently examined variables in the prediction of aggression (for a meta-analytic review, see de Castro et al., 2002). A hostile attribution bias can affect SIP by guiding the interpretation of social cues (SIP step 2) in a way that the probability to show aggressive behavior is increased. For the appearance-related domain, it is most likely that there are indirect rather than direct relations with aggression; for example, by increasing the probability of specific social experiences that are stored in the database and that, in turn, may influence the development of social behavior. Besides physical attractiveness, ethnicity can play a very influential role in this context. A study by Bellmore, Nishina, Witkow, Graham, and Juvonen (2007) was able to show that there is a same-ethnicity bias when it comes to peer relationships in middle childhood.

Peers of the same ethnicity received greater acceptance and less rejection compared to peers of another ethnicity. As in Western cultures the vast majority of people is white (e.g., 72.4% in 2010 in the United States, Humes, Jones, & Ramirez, 2011), children of another ethnicity (and corresponding appearance) could be at a higher risk to experience peer problems, which in turn influences the database, and, thus, the development of aggressive behavior.

Even though the presented projects were focused on intrapersonal risk factors of aggressive behavior, it is not implied that other risk factors, interpersonal and environmental, are not important. In fact, past research revealed numerous of such factors that can be predictive of aggression. For instance, as the results in Paper 3 suggest problems in peer relationships are correlated with the development of aggressive behavior (for a meta-analytic review, see Reijntjes et al., 2011). Another example would be emotional maltreatment and physical abuse in infancy and early childhood that is related to externalizing behavior (Manly et al., 2001).

In addition, the intrapersonal factors that were in the scope of the presented empirical papers may partially be influenced by interpersonal factors. For example, a study by Otterpohl et al. (2012) revealed that children's ability to regulate their angry feelings can be predicted by parental practices. In particular, parental punitive reactions in response to children's anger were correlated with the use of maladaptive regulation strategies. Regarding theory of mind, a study by Hughes and Dunn (1998) showed that mental-state talk with peer friends can be beneficial. Three-year-old children who talked more about mental states with their friends compared to those who talked less about mental states scored higher in theory of mind tasks on a subsequent occasion. Regarding physical attractiveness, the presented study in Paper 3 itself suggests that the quality of peer relationships can be predictive of later attractiveness. Children who experienced more problems with their peers were rated as less physically attractive by independent adult raters on a later occasion than children who experienced less peer problems (stress may be a mediating variable for this link, as discussed in Paper 3).

6.3 General Strengths and Limitations

This doctoral dissertation and its included empirical papers have several strengths and limitations that are noteworthy. A first strength is the large sample size of N = 1,657 children in the age of 6 to 11 years at T1 in combination with the quite low attrition rate of 9.5% from T1 to T3 (see Chapter 4.1). Even though the total sample was only used for the analyses in Paper 2, the subsamples drawn for Paper 1 and 3 can also be considered large when taking into account the context of behavioral observation (N = 599) and attractiveness research (N = 1,043),

respectively. Such large sample sizes are beneficial as sample size is linked to statistical power (e.g., Field, 2009). Statistical power is the probability to find an effect or correlation when one or the other is actually present. Thus, high power is important, in particular in longitudinal autoregressive models, where effect sizes are often quite small (Adachi & Willoughby, 2015). Indeed, the main effect sizes in the presented studies were small (e.g., $\beta = -.06$ for the link of attractiveness T1 to peer problems T2 in Paper 3), but for the most part significant at $p < .05$ due to high statistical power.

Another strength is the longitudinal format of the empirical studies and the fact that the central constructs were measured not only at one single measurement point. Specifically, the risk factors were assessed at two measurement points each (anger regulation at T1 and T3, theory of mind at T1 and T2, physical attractiveness at T1 and T3) and the teacher-report of aggressive behavior, the core element of all central analyses, was conducted at all three measurement points. This allowed to control for past levels of the outcome (i.e., temporal stability) in all central analyses. In addition, all central analyses considered initial correlations between the risk factors and aggression, and potential reciprocal effects. The resulting cross-lagged panel-like models foster (but not prove) a causal interpretation of the found effects (Marmor & Montemayor, 1977).

As a final methodological strength, the use of different informants to assess the constructs is noteworthy. Specifically, all central analyses used the teacher-report of aggressive behavior, while the risk factors were measured using another informant than the teachers. Anger regulation was assessed by a behavioral observation of children's actual behavior in an anger-eliciting situation, theory of mind was assessed by children's responses to cartoons, and physical attractiveness was assessed by independent adult raters based on facial photos of the children. That is, there was no informant overlap between the measure of aggression and the measure of the corresponding risk factor in either central analysis. This feature has the advantage to reduce common method biases, a common problem in behavioral research that is linked to systematic measurement errors (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003).

Despite these strengths, there are some limitations that should be acknowledged. The biggest limitation of the presented studies is the fact that the SIP model, which functions as a framework of this doctoral dissertation, was not explicitly tested. Based on the SIP model by Crick and Dodge (1994), it was proposed that the examined risk factors have an impact on SIP, which in turn influences the development of aggressive behavior in the long run. For example, it was hypothesized that maladaptive anger regulation leads to maintained or even increased

levels of anger that may guide the interpretation of social cues (e.g., hostile attributions; SIP step 2) and the setting of goals (e.g., harmful goals; SIP step 3). In a similar manner, it was assumed that deficits in theory of mind are associated with biased SIP as the lack of understanding others' behavior may also lead to incorrect interpretation of social cues (e.g., hostile attributions; SIP step 2). For physical attractiveness, it was proposed that unfavorable social experiences (i.e., peer problems) result from low attractiveness and are stored in the database. The database, in turn, influences virtually all SIP steps. For example, the associated worse outcome expectations in a social encounter may lead to less beneficial response decisions in the same encounter (SIP step 5). In this sense, it was assumed that all examined risk factors indirectly predict aggression through biased SIP patterns. However, this assumption was not tested as there were no measures of SIP patterns in the PIER-Project. Instead, all central analyses can be seen as "blackbox tests": The risk factors were the input, aggression was the output, but the underlying mechanisms remain unclear in the presented studies (with an exception of the factor attractiveness; here, peer problems could be identified as an underlying mechanism). Thus, future studies could replicate the findings provided in this work, and extent them by adding SIP-specific variables, such as hostile versus nonhostile attributions of intent or the generation of aggressive versus nonaggressive response options (e.g., Lansford et al., 2010).

Another limitation lies in the assessment of aggressive behavior. Aggression was measured via teacher-reports, children's self-reports, and a behavioral measure (see Chapter 4.3.1). Even though this is a variety of aggression measures, only the teacher-report was considered to be appropriate for the longitudinal analyses. The reason was that only this measure was assessed at all three measurement points. The behavioral measure was only assessed at T3, and the self-report only at T2 and T3. In addition, the use of self-reports on aggression may be problematic as aggressive behavior is obviously socially undesirable and, thus, the resulting social desirability responding may lead to an underreporting of aggression (e.g., Vigil-Colet et al., 2012). This is especially problematic considering the fact that at T1 and T2 the children gave their responses not privately, but by pointing at the corresponding response option (see Chapter 4.2). Admittedly, teacher-reports also may be problematic to some extents. For example, teachers may be unaware of more covert forms of aggression, such as spreading rumors (i.e., children's engagement in relational aggression). Therefore, future studies could make an effort to use more ecologically valid measures of aggressive behavior and include these measures from the first measurement point. One of such alternative measures could be peer

nominations; that is, children nominate classmates that behave in an aggressive way (e.g., Kokko, Pulkkinen, Huesmann, Dubow, & Boxer, 2009).

6.4 Implications

Based on the obtained results in the presented studies and the discussion in the doctoral dissertation, several implications can be inferred, both for future studies and for preventing aggressive behavior in children. On a more general level, the results are for the most part in line with the frequently cited SIP model (Crick & Dodge, 1994), but explicit tests for potential underlying SIP mechanisms are still missing, as discussed above. Thus, future studies should replicate the presented findings and extent them by examining potential underlying SIP variables.

The methodological study presented in Paper 1 provides a reliable, valid, and easy to use behavioral observation approach to assess actual anger regulation strategy use in childhood. The development of such measures is of scientific relevance as the often used self-reports are fraught with diverse problems (Underwood, 1997). In combination with previous publications examining the validity of the original form of the behavioral observation assessment (Rohlf & Krahé, 2015; Kirsch et al., 2015), future research can resort to two variations of this measure that are conceptually equivalent, but sufficiently different for the use in longitudinal studies, and for the use in different age ranges (6- to 10-year olds vs. 9- to 13-year olds). However, although this assessment overcomes most of the problems attached to self-reports, it also has some limitations. For example, behavioral observation can only capture anger regulation strategies that are accessible to observation (e.g., mental strategies are not assessed). Thus, future studies should make an effort to choose the appropriate tool to answer their research questions. In this context, it is important to consider children's age, the context in which emotions are induced (e.g., social vs. non-social), and the developmental significance of internal versus external regulation. For example, as they get older, children's use of mental strategies gains in importance (e.g., McRae et al., 2012).

On a practical level, the findings can be fruitful for the conceptualization of programs that are aimed to prevent or reduce aggressive behavior in children (for a review, see Gollwitzer et al., 2007; Wilson & Lipsey, 2007). For example, theory of mind was identified as a cognitive risk factor of aggression in Paper 2. Thus, explicitly enhancing theory of mind skills could be an important feature of such programs. In fact, some aspects of theory of mind are trained in prominent prevention programs, such as identifying own and others' feelings in the "I can

Problem Solve" program (ICPS) that was found to reduce aggressive behavior in children (Boyle & Hassett-Walker, 2008). As discussed above, a high frequency of mental-state talks is predictive of superior theory of mind (Hughes & Dunn, 1998). Thus, programs additionally could encourage children to talk about their or others' mental states with peers or parents in order to further improve theory of mind. As children become more skilled in comprehending others' behavior and intentions, biases in SIP are diminishing, which in turn results in a reduced probability to show aggressive behavior. As the results from Paper 2 suggest, the effects could be expected for both the physical and the relational form of aggression.

Regarding low physical attractiveness as a risk factor of aggression, it should not be concluded that beautifying all children is a suggested implication. The results from Paper 3 emphasize the importance of experienced peer problems in the link between low physical attractiveness and aggressive behavior. However, as discussed above, it remains unclear whether unattractive children experienced more problematic peer relationships caused by a genuine correlation between attractiveness and social skills, or by generated unfavorable, but incorrect expectations of others (i.e., stereotype-based). Under the latter assumption, the goal should be the reduction of discrimination based on appearance. In particular, programs could aim to reduce appearance-based stereotypes in order to reduce social discrimination and, thus, aggressive behavior. Promoting media literacy in children may be one important starting point to challenge these stereotypes as media shapes a consumer's values and conception of the world, according to cultivation theory (for a review, see Gerbner, Gross, Morgan, Signorielli, & Shanahan, 2002). A constant use of media is natural these days. For example, 8- to 12-year-old children in the United States spend about six hours every day using media for entertainment, including playing video games, watching movies, or using social media (Common Sense, 2015). At the same time, these media very often convey the impression that attractiveness equals a positive personality; that is, media promote the attractiveness stereotype. For example, Bazzini, Curtin, Joslin, Regan, and Martz (2010) revealed that even in Disney movies, physically attractive characters in comparison to unattractive characters are portrayed in a more social desirable way, including that they are displayed as low aggressive, but very friendly and intelligent. This may also be the case in modern media. Nowadays, children and adolescents prefer consuming online videos, in particular YouTube, over traditional television (DEFY Media, 2015). In this medium, the topic "beauty" tremendously gained in attention within the last few years (Pixability, 2014). At the same time, although content on YouTube is predominantly produced by ordinary citizens rather than professional organizations, stereotypes usually are reinforced rather than challenged (e.g., Guo & Harlow, 2014). Teaching the children

in questioning media-communicated stereotypes proactively by analyzing and evaluating messages sent by media (media literacy) could be beneficial in reducing the negative social consequences less attractive individuals have to face. Another set of interventions may start in the classroom. In his review on reducing stereotype threat, Sparks (2016) summarized diverse of such approaches, including the presentation of positive role models, supporting students' sense of belonging, and valuing students' individuality. As the results in Paper 3 suggest, it can be expected that the indirect link between attractiveness and aggression is diminished if the direct link between low attractiveness and peer problems is decreased. Finally, these practical implications not only apply to the indirect effect physical attractiveness can have on an individual's development, but also to all kinds of social stereotypes that could have the potential for a negative impact, such as ethnicity or sexual orientation.

The presented practical implications should be evaluated in light of the effect sizes obtained in the empirical studies. Admittedly, the effects were quite small for both theory of mind (bivariate correlations with aggression at T3: *r*s ranging from -.08 to -.12) and physical attractiveness (bivariate correlation with aggression T3: $r = -.11$). However, these magnitudes are comparable to other well-known intrapersonal risk factors, such as the often cited hostile attribution bias, which is correlated with $r = .13$ with aggressive behavior in a meta-analysis by de Castro et al. (2002; regarding nonreferred general samples of children and adolescents). In addition, it should be considered that the prediction of aggressive behavior is a complex issue and there are many factors identified so far that contribute to the occurrence and development of aggression (for a review, see Krahé, 2013; Valois et al., 2002). Thus, it can be stated that, alongside other important developmental factors, theory of mind and physical attractiveness can have a substantial impact on the development of aggressive behavior, and, consequently, provide points to prevent aggressive behavior in childhood.

In conclusion, this doctoral dissertation expands past research on risk factors of aggressive behavior and emphasizes the complexity of predicting aggressive behavior. Considering the numerous risk factors identified in the current and previous research, it can be stated that the development of aggression is clearly based on a multifactorial system.

References

Adachi, P., & Willoughby, T. (2015). Interpreting effect sizes when controlling for stability effects in longitudinal autoregressive models: Implications for psychological science. *European Journal of Developmental Psychology*, *12*, 116–128. doi:10.1080/17405629.2014.963549

Adrian, M., Zeman, J., & Veits, G. (2011). Methodological implications of the affect revolution: A 35-year review of emotion regulation assessment in children. *Journal of Experimental Child Psychology*, *110*, 171–197. doi:10.1016/j.jecp.2011.03.009

Agthe, M., Aydin, N., Pfundmair, M., Frey, D., & Niesta Kayser, D. (2016). Physische Attraktivität. In H.-W. Bierhoff & D. Frey (Eds.), *Enzyklopädie der Psychologie. Soziale Motive und soziale Einstellungen* [Encyclopedia of psychology. Social motives and social attitudes] (pp. 298–315). Göttingen, D: Hogrefe.

Aldao, A., Nolen-Hoeksema, S., & Schweizer, S. (2010). Emotion-regulation strategies across psychopathology: A meta-analytic review. *Clinical Psychology Review*, *30*, 217–237. doi:10.1016/j.cpr.2009.11.004

Alink, L. R. A., Mesman, J., van Zeijl, J., Stolk, M. N., Juffer, F., Koot, H. M.,. . . van Ijzendoorn, M. H. (2006). The early childhood aggression curve: Development of physical aggression in 10- to 50-month-old children. *Child Development*, *77*, 954–966. doi:10.1111/j.1467-8624.2006.00912.x

Altshuler, J. L., & Ruble, D. N. (1989). Developmental changes in children's awareness of strategies for coping with uncontrollable stress. *Child Development*, *60*, 1337–1349. doi:10.1111/1467-8624.ep9772243

Archer, J. (2004). Sex differences in aggression in real-world settings: A meta-analytic review. *Review of General Psychology*, *8*, 291–322. doi:10.1037/1089-2680.8.4.291

Archer, J., & Coyne, S. M. (2005). An integrated review of indirect, relational, and social aggression. *Personality and Social Psychology Review*, *9*, 212–230. doi:10.1207/s15327957pspr0903_2

Asendorpf, J. B., & van Aken, M. A. G. (1993). Deutsche Versionen der Selbstkonzeptskalen von Harter [German versions of Harter's self-concept scales for children]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, *25*, 64–86.

Asparouhov, T., & Muthén, B. (2010a). Bayesian analysis of latent variable models using Mplus. Retrieved from http://www.statmodel.com/download/BayesAdvantages18.pdf

Asparouhov, T., & Muthén, B. (2010b). Chi-square statistics with multiple imputation. Retrieved from http://statmodel.com/download/MI7.pdf

Asparouhov, T., & Muthén, B. (2010c). Weighted least squares estimation with missing data. Retrieved from https://www.statmodel.com/download/GstrucMissingRevision.pdf

Baron, R. A., & Richardson, D. S. (1994). *Human aggression* (2nd ed.). New York, NY, US: Plenum Press.

Bartolotta, T. E., & Shulman, B. B. (2010). Child development. In B. B. Shulman & N. C. Singleton (Eds.), *Language development. Foundations, processes, and clinical applications* (pp. 35–53). Sudbury, MA, US: Jones and Bartlett Publishers.

Bazzini, D., Curtin, L., Joslin, S., Regan, S., & Martz, D. (2010). Do animated Disney characters portray and promote the Beauty-Goodness Stereotype? *Journal of Applied Social Psychology*, *40*, 2687–2709. doi:10.1111/j.1559-1816.2010.00676.x

Bechara, A. (2004). The role of emotion in decision-making: Evidence from neurological patients with orbitofrontal damage. *Brain and Cognition*, *55*, 30–40. doi:10.1016/j.bandc.2003.04.001

Bellmore, A. D., Nishina, A., Witkow, M. R., Graham, S., & Juvonen, J. (2007). The influence of classroom ethnic composition on same- and other-ethnicity peer nominations in middle school. *Social Development*, *16*, 720–740. doi:10.1111/j.1467-9507.2007.00404.x

Benzeval, M., Green, M. J., & Macintyre, S. (2013). Does perceived physical attractiveness in adolescence predict better socioeconomic position in adulthood? Evidence from 20 years of follow up in a population cohort study. *PLOS ONE*, *8*, e63975. doi:10.1371/journal.pone.0063975

Bergen, D. (2002). The role of pretend play in children's cognitive development. *Early Childhood Research and Practice*, *4,* 1-12.

Berkowitz, L., & Frodi, A. (1979). Reactions to a child's mistakes as affected by her/his looks and speech. *Social Psychology Quarterly*, *42*, 420–425. doi:10.2307/3033814

Bobadilla, L., Metze, A. V., & Taylor, J. (2013). Physical attractiveness and its relation to unprovoked and reactive aggression. *Journal of Research in Personality*, *47*, 70–77. doi:10.1016/j.jrp.2012.09.004

Bonica, C., Arnold, D. H., Fisher, P. H., Zeljo, A., & Yershova, K. (2003). Relational aggression, relational victimization, and language development in preschoolers. *Social Development*, *12*, 551–562. doi:10.1111/1467-9507.00248

Borch, C., Hyde, A., & Cillessen, A. H. N. (2011). The role of attractiveness and aggression in high school popularity. *Social Psychology of Education*, *14*, 23–39. doi:10.1007/s11218-010-9131-1

Boyle, D., & Hassett-Walker, C. (2008). Reducing overt and relational aggression among young children. *Journal of School Violence*, *7*, 27–42. doi:10.1300/J202v07n01_03

Brenner, E. M., & Salovey, P. (1997). Emotion regulation during childhood: Developmental, interpersonal, and individual considerations. In P. Salovey & D. J. Sluyter (Eds.), *Emotional development and emotional intelligence. Educational implications* (1st ed., pp. 168–195). New York, NY, US: Basic Books.

Calvete, E., & Orue, I. (2012). The role of emotion regulation in the predictive association between social information processing and aggressive behavior in adolescents. *International Journal of Behavioral Development*, *36*, 338–347. doi:10.1177/0165025412444079

Caputi, M., Lecce, S., Pagnin, A., & Banerjee, R. (2012). Longitudinal effects of theory of mind on later peer relations: The role of prosocial behavior. *Developmental Psychology*, *48*, 257–270. doi:10.1037/a0025402

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. doi:10.1111/j.1467-8624.2008.01184.x

Carlson, J. M., Dikecligil, G. N., Greenberg, T., & Mujica-Parodi, L. R. (2012). Trait reappraisal is associated with resilience to acute psychological stress. *Journal of Research in Personality*, *46*, 609–613. doi:10.1016/j.jrp.2012.05.003

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*, 464–504. doi:10.1080/10705510701301834

Choe, D. E., Lane, J. D., Grabell, A. S., & Olson, S. L. (2013). Developmental precursors of young school-age children's hostile attribution bias. *Developmental Psychology*, *49*, 2245–2256. doi:10.1037/a0032293

Cole, P. M., & Hall, S. E. (2008). Emotion dysregulation as a risk factor for psychopathology. In T. P. Beauchaine & S. P. Hinshaw (Eds.), *Child and adolescent psychopathology* (pp. 265–298). Hoboken, NJ, US: John Wiley & Sons.

Cole, P. M., Michel, M. K., & Teti, L. O. (1994). The development of emotion regulation and dysregulation: A clinical perspective. *Monographs of the Society for Research in Child Development*, *59*, 73–100. doi:10.2307/1166139

Collins, W. A. (Ed.). (1984). *Development during middle childhood: The years from six to twelve*. Washington, DC, US: National Academy Press.

Common Sense. (2015). The common sense census: Media use by tweens and teens. Retrieved from https://www.commonsensemedia.org/sites/default/files/uploads/research/census_researchreport.pdf

Crick, N. R. (1996). The role of overt aggression, relational aggression, and prosocial behavior in the prediction of children's future social adjustment. *Child Development*, *67*, 2317–2327. doi:10.2307/1131625

Crick, N. R. (1997). Engagement in gender normative versus nonnormative forms of aggression: Links to social-psychological adjustment. *Developmental Psychology*, *33*, 610–617. doi:10.1037/0012-1649.33.4.610

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. doi:10.1037/0033-2909.115.1.74

Crick, N. R., Werner, N. E., Casas, J. F., O'Brien, K. M., Nelson, D. A., Grotpeter, J. K., & Markon, K. (1999). Childhood aggression and gender: A new look at an old problem. In D. Bernstein (Ed.), *Gender and motivation* (pp. 75–141). Lincoln, NE, US: University of Nebraska Press.

Crockenberg, S. C., & Leerkes, E. M. (2004). Infant and maternal behaviors regulate infant reactivity to novelty at 6 months. *Developmental Psychology*, *40*, 1123–1132. doi:10.1037/0012-1649.40.6.1123

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281–302. doi:10.1037/h0040957

Cummings, E. M., Davies, P., & Campbell, S. B. (2000). *Developmental psychopathology and family process: Theory, research, and clinical implications*. New York, NY, US: Guilford Press.

Darley, J. M., & Fazio, R. H. (1980). Expectancy confirmation processes arising in the social interaction sequence. *American Psychologist*, *35*, 867–881. doi:10.1037/0003-066X.35.10.867

de Castro, B. O., Merk, W., Koops, W., Veerman, J. W., & Bosch, J. D. (2005). Emotions in social information processing and their relations with reactive and proactive aggression in referred aggressive boys. *Journal of Clinical Child and Adolescent Psychology*, *34*, 105–116. doi:10.1207/s15374424jccp3401_10

de Castro, B. O., Veerman, J. W., Koops, W., Bosch, J. D., & Monshouwer, H. J. (2002). Hostile attribution of intent and aggressive behavior: A meta-analysis. *Child Development*, *73*, 916–934. doi:10.1111/1467-8624.00447

DEFY Media. (2015). Acumen report: Constant content. Retrieved from http://cdn.defymedia.com/wp-content/uploads/2015/10/Acumen-Report-Constant-Content.pdf

Denson, T. F., Moulds, M. L., & Grisham, J. R. (2012). The effects of analytical rumination, reappraisal, and distraction on anger experience. *Behavior Therapy*, *43*, 355–364. doi:10.1016/j.beth.2011.08.001

Devine, R. T., White, N., Ensor, R., & Hughes, C. (2016). Theory of mind in middle childhood: Longitudinal associations with executive function and social competence. *Developmental Psychology*, *52*, 758–771. doi:10.1037/dev0000105

Dimitrova, V., & Lüdmann, M. (2014). *Sozial-emotionale Kompetenzentwicklung: Leitlinien der Entfaltung der emotionalen Welt* [Social-emotional competence development: Guidelines for the development of the emotional world]. Wiesbaden, D: Springer VS.

Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology*, *24*, 285–290. doi:10.1037/h0033731

Dion, K. K. (1973). Young children's stereotyping of facial attractiveness. *Developmental Psychology*, *9*, 183–188. doi:10.1037/h0035083

Dodge, K. A., Lansford, J. E., Burks, V. S., Bates, J. E., Pettit, G. S., Fontaine, R., & Price, J. M. (2003). Peer rejection and social information-processing factors in the development of aggressive behavior problems in children. *Child Development*, *74*, 374–393. doi:10.1111/1467-8624.7402004

Dumas, J. E., Nilsen, W., & Lynch, A. M. (2001). How much does physical appearance say about the psychological adjustment of competent and dysfunctional children? *Journal of Clinical Child Psychology*, *30*, 385–398. doi:10.1207/S15374424JCCP3003_10

Dunkake, I., Kiechle, T., Klein, M., & Rosar, U. (2012). Schöne schüler, schöne Noten? Eine empirische Untersuchung zum Einfluss der physischen Attraktivität von Schülern auf die Notenvergabe durch das Lehrpersonal [Good looks, good grades? An empirical analysis of the influence of students' physical attractiveness on grading by teachers]. *Zeitschrift für Soziologie*, *41*, 142–161.

Eagly, A. H., Ashmore, R. D., Makhijani, M. G., & Longo, L. C. (1991). What is beautiful is good, but…: A meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin*, *110*, 109–128. doi:10.1037/0033-2909.110.1.109

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

Eisenberg, N., & Fabes, R. A. (1995). The role of emotionality and regulation in children's social functioning: A longitudinal study. *Child Development*, *66*, 1360–1384. doi:10.1111/1467-8624.ep9510075268

Eisenthal, Y., Dror, G., & Ruppin, E. (2006). Facial attractiveness: Beauty and the machine. *Neural Computation*, *18*, 119–142. doi:10.1162/089976606774841602

Ekas, N. V., Lickenbrock, D. M., & Braungart-Rieker, J. M. (2013). Developmental trajectories of emotion regulation across infancy: Do age and the social partner influence temporal patterns? *Infancy*, *18*, 729–754. doi:10.1111/infa.12003

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, *17*, 124–129. doi:10.1037/h0030377

Enders, C., & Bandalos, D. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling*, *8*, 430–457. doi:10.1207/S15328007SEM0803_5

English, T., Lee, I. A., John, O. P., & Gross, J. J. (2017). Emotion regulation strategy selection in daily life: The role of social context and goals. *Motivation and Emotion*, *41*, 230–242. doi:10.1007/s11031-016-9597-z

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). Los Angeles, CA, US: Sage.

Forgas, J. P. (2011). She just doesn't look like a philosopher…? Affective influences on the halo effect in impression formation. *European Journal of Social Psychology*, *41*, 812–817. doi:10.1002/ejsp.842

Frith, U., & Frith, C. (2010). The social brain: Allowing humans to boldly go where no other species has been. *Philosophical Transactions of the Royal Society of London*, *365*, 165–176. doi:10.1098/rstb.2009.0160

Gerbner, G., Gross, L., Morgan, M., Signorielli, N., & Shanahan, J. (2002). Growing up with television: Cultivation processes. In J. Bryant & D. Zillmann (Eds.), *Media effects. Advances in theory and research* (2nd ed., pp. 43–67). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Glocker, M. L., Langleben, D. D., Ruparel, K., Loughead, J. W., Gur, R. C., & Sachser, N. (2009). Baby schema in infant faces induces cuteness perception and motivation for caretaking in adults. *Ethology*, *115*, 257–263. doi:10.1111/j.1439-0310.2008.01603.x

Gollwitzer, M., Pfetsch, J., Schneider, V., Schulz, A., Steffke, T., & Ulrich, C. (Eds.). (2007). *Gewaltprävention bei Kindern und Jugendlichen* [Violence prevention for children and adolescents]. Göttingen, D: Hogrefe.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry*, *38*, 581–586. doi:10.1111/j.1469-7610.1997.tb01545.x

Griffin, A. M., & Langlois, J. H. (2006). Stereotype directionality and attractiveness stereotyping: Is beauty good or is ugly bad? *Social Cognition*, *24*, 187–206. doi:10.1521/soco.2006.24.2.187

Grob, A., & Smolenski, C. (2005). *Fragebogen zur Erhebung der Emotionsregulation bei Kindern und Jugendlichen (FEEL-KJ)* [Questionnaire on emotion regulation in children and adolescents (FEEL-KJ)]. Bern, CH: Verlag Hans Huber.

Gross, J. J. (1998). The emerging field of emotion regulation: An integrative review. *Review of General Psychology*, *2*, 271–299. doi:10.1037/1089-2680.2.3.271

Gross, J. J. (2002). Emotion regulation: Affective, cognitive, and social consequences. *Psychophysiology*, *39*, 281–291. doi:10.1017/S0048577201393198

Gross, J. J., Carstensen, L. L., Pasupathi, M., Tsai, J., Götestam Skorpen, C., & Hsu, A. Y. C. (1997). Emotion and aging: Experience, expression, and control. *Psychology and Aging*, *12*, 590–599. doi:10.1037/0882-7974.12.4.590

Gross, J. J., & Thompson, R. A. (2007). Emotion regulation: Conceptual foundations. In J. J. Gross (Ed.), *Handbook of emotion regulation* (pp. 3–24). New York, NY, US: Guilford Press.

Grueneisen, S., Wyman, E., & Tomasello, M. (2015). "I know you don't know I know…" children use second-order false-belief reasoning for peer coordination. *Child Development*, *86*, 287–293. doi:10.1111/cdev.12264

Guo, L., & Harlow, S. (2014). User-generated racism: An analysis of stereotypes of African Americans, Latinos, and Asians in YouTube videos. *Howard Journal of Communications*, *25*, 281–302. doi:10.1080/10646175.2014.925413

Hahn, A. C., & Perrett, D. I. (2014). Neural and behavioral responses to attractiveness in adult and infant faces. *Neuroscience and Biobehavioral Reviews*, *46*, 591–603. doi:10.1016/j.neubiorev.2014.08.015

Hayes, A. F., & Scharkow, M. (2013). The relative trustworthiness of inferential tests of the indirect effect in statistical mediation analysis: Does method really matter? *Psychological Science*, *24*, 1918–1927. doi:10.1177/0956797613480187

Helmsen, J., & Petermann, F. (2010). Emotionsregulationsstrategien und aggressives Verhalten im Kindergartenalter [Emotion regulation strategies and aggressive behavior of preschool children]. *Praxis der Kinderpsychologie und Kinderpsychiatrie*, *59*, 775–791.

Henry, J. D., Phillips, L. H., Ruffman, T., & Bailey, P. E. (2013). A meta-analytic review of age differences in theory of mind. *Psychology and Aging*, *28*, 826–839. doi:10.1037/a0030677

Hermann, J. M., & Vollmeyer, R. (2016). Stereotype Threat in der Grundschule [Stereotype threat in primary school]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, *48*, 42–49. doi:10.1026/0049-863 7/a000143

Hildebrandt, K. A., & Fitzgerald, H. E. (1979). Facial feature determinants of perceived infant attractiveness. *Infant Behavior and Development*, *2*, 329–339. doi:10.1016/S0163-6383(79)80043-0

Huesmann, L. R., Dubow, E. F., & Boxer, P. (2009). Continuity of aggression from childhood to early adulthood as a predictor of life outcomes: Implications for the adolescent-limited and life-course-persistent models. *Aggressive Behavior*, *35*, 136–149. doi:10.1002/ab.20300

Hughes, C., & Dunn, J. (1998). Understanding mind and emotion: Longitudinal associations with mental-state talk between young friends. *Developmental Psychology*, *34*, 1026–1037. doi:10.1037/0012-1649.34.5.1026

Hughes, C., & Leekam, S. (2004). What are the links between theory of mind and social relations? Review, reflections and new directions for studies of typical and atypical development. *Social Development*, *13*, 590–619. doi:10.1111/j.1467-9507.2004.00285.x

Humes, K. R., Jones, N. A., & Ramirez, R. R. (2011). *Overview of Race and Hispanic Origin: 2010*. Washington, DC, US: U.S. Census Bureau.

Inzlicht, M., & Schmader, T. (Eds.). (2012). *Stereotype threat: Theory, process, and application*. Oxford, UK: Oxford University Press.

Jung, J., Krahé, B., Bondü, R., Esser, G., & Wyschkon, A. (2016). Dynamic progression of antisocial behavior in childhood and adolescence: A three-wave longitudinal study from Germany. *Applied Developmental Science*, *22*, 1–15. doi:10.1080/10888691.2016.1219228

Kalbe, E., Schlegel, M., Sack, A. T., Nowak, D. A., Dafotakis, M., Bangard, C.,. . . Kessler, J. (2010). Dissociating cognitive from affective theory of mind: A TMS study. *Cortex*, *46*, 769–780. doi:10.1016/j.cortex.2009.07.010

Kanazawa, S., & Kovar, J. L. (2004). Why beautiful people are more intelligent. *Intelligence*, *32*, 227–243. doi:10.1016/j.intell.2004.03.003

Kirsch, F., Rohlf, H., & Krahé, B. (2015). Measuring anger regulation in middle childhood through behavioural observation: A longitudinal validation. *European Journal of Developmental Psychology*, *12*, 718–727. doi:10.1080/17405629.2015.1101375

Kokkinos, C. M., Kakarani, S., & Kolovou, D. (2016). Relationships among shyness, social competence, peer relations, and theory of mind among pre-adolescents. *Social Psychology of Education*, *19*, 117–133. doi:10.1007/s11218-015-9317-7

Kokkinos, C. M., Voulgaridou, I., Mandrali, M., & Parousidou, C. (2016). Interactive links between relational aggression, theory of mind, and moral disengagement among early adolescents. *Psychology in the Schools*, *53*, 253–269. doi:10.1002/pits.21902

Kokko, K., Pulkkinen, L., Huesmann, L. R., Dubow, E. F., & Boxer, P. (2009). Intensity of aggression in childhood as a predictor of different forms of adult aggression: A two-country (Finland and United States) analysis. *Journal of Research on Adolescence*, *19*, 9–34. doi:10.1111/j.1532-7795.2009.00579.x

Kopp, C. B. (1992). Emotional distress and control in young children. *New Directions for Child and Adolescent Development*, *1992*, 41–56. doi:10.1002/cd.23219925505

Krahé, B. (2013). *The social psychology of aggression* (2nd ed.). Hove, UK: Psychology Press.

Krahé, B. (2016). Violent media effects on aggression: A commentary from a cross-cultural perspective. *Analyses of Social Issues and Public Policy*, *16*, 439–442. doi:10.1111/asap.12107

Kun, B., & Demetrovics, Z. (2010). Emotional intelligence and addictions: A systematic review. *Substance Use & Misuse*, *45*, 1131–1160. doi:10.3109/10826080903567855

Lam, C. B., McHale, S. M., & Crouter, A. C. (2014). Time with peers from middle childhood to late adolescence: Developmental course and adjustment correlates. *Child Development*, *85*, 1677–1693. doi:10.1111/cdev.12235

Langlois, J. H., & Downs, A. C. (1979). Peer relations as a function of physical attractiveness: The eye of the beholder or behavioral reality? *Child Development*, *50*, 409–418. doi:10.1111/j.1467-8624.1979.tb04122.x

Langlois, J. H., Kalakanis, L., Rubenstein, A. J., Larson, A., Hallam, M., & Smoot, M. (2000). Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychological Bulletin*, *126*, 390–423. doi:10.1037//0033-2909.126.3.390

Langlois, J. H., Ritter, J. M., Casey, R. J., & Sawin, D. B. (1995). Infant attractiveness predicts maternal behaviors and attitudes. *Developmental Psychology*, *31*, 464–472. doi:10.10 37/0012-1649.31.3.464

Lansford, J. E., Malone, P. S., Dodge, K. A., Pettit, G. S., & Bates, J. E. (2010). Developmental cascades of peer rejection, social information processing biases, and aggression during middle childhood. *Development and Psychopathology*, *22*, 593–602. doi:10.1017/S0954579 410000301

Lansford, J. E., Skinner, A. T., Sorbring, E., Di Giunta, L., Deater-Deckard, K., Dodge, K. A.,. . . Chang, L. (2012). Boys' and girls' relational and physical aggression in nine countries. *Aggressive Behavior*, *38*, 298–308. doi:10.1002/ab.21433

Leary, M. R., Twenge, J. M., & Quinlivan, E. (2006). Interpersonal rejection as a determinant of anger and aggression. *Personality and Social Psychology Review*, *10*, 111–132. doi:10.12 07/s15327957pspr1002_2

Lemerise, E. A., & Arsenio, W. F. (2000). An integrated model of emotion processes and cognition in social information processing. *Child Development*, *71*, 107–118. doi:10.11 11/1467-8624.00124

Li, C.-H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods*, *48*, 936–949. doi:10.3758/s13428-015-0619-7

Manly, J. T., Kim, J. E., Rogosch, F. A., & Cicchetti, D. (2001). Dimensions of child maltreatment and children's adjustment: Contributions of developmental timing and subtype. *Development and Psychopathology*, *13*, 759-782.

Marcovitch, S., O'Brien, M., Calkins, S. D., Leerkes, E. M., Weaver, J. M., & Levine, D. W. (2015). A longitudinal assessment of the relation between executive function and theory of mind at 3, 4, and 5 years. *Cognitive Development*, *33*, 40–55. doi:10.1016/j.cogdev.20 14.07.001

Marmor, G. S., & Montemayor, R. (1977). The cross-lagged panel design: A review. *Perceptual and Motor Skills*, *45*, 883–893. doi:10.2466/pms.1977.45.3.883

McLaughlin, K. A., Hatzenbuehler, M. L., Mennin, D. S., & Nolen-Hoeksema, S. (2011). Emotion dysregulation and adolescent psychopathology: A prospective study. *Behaviour Research and Therapy*, *49*, 544–554. doi:10.1016/j.brat.2011.06.003

McRae, K., Gross, J. J., Weber, J., Robertson, E. R., Sokol-Hessner, P., Ray, R. D.,. . . Ochsner, K. N. (2012). The development of emotion regulation: An fMRI study of cognitive reappraisal in children, adolescents and young adults. *Social Cognitive and Affective Neuroscience*, *7*, 11–22. doi:10.1093/scan/nsr093

Mehrabian, A., & Blum, J. S. (1997). Physical appearance, attractiveness, and the mediating role of emotions. *Current Psychology*, *16*, 20–42. doi:10.1007/s12144-997-1013-0

Miller, S. A. (2009). Children's understanding of second-order mental states. *Psychological Bulletin*, *135*, 749–773. doi:10.1037/a0016854

Moffitt, T. E. (1993). Adolescence-limited and life-course-persistent antisocial behavior: A developmental taxonomy. *Psychological Review*, *100*, 674–701. doi:10.1037/0033-295X.100.4.674

Moffitt, T. E. (2007). A review of research on the taxonomy of life-course persistent versus adolescence-limited antisocial behavior. In D. J. Flannery, A. T. Vazsonyi, & I. D. Waldman (Eds.), *The Cambridge handbook of violent behavior and aggression* (pp. 49–74). Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511816840.004

Murray-Close, D., Ostrov, J. M., & Crick, N. R. (2007). A short-term longitudinal study of growth of relational aggression during middle childhood: Associations with gender, friendship intimacy, and internalizing problems. *Development and Psychopathology*, *19*, 187–203. doi:10.1017/S0954579407070101

Muthén, L. K., & Muthén, B. O. (1998-2015). *Mplus user's guide* (7th ed.). Los Angeles, CA, US: Muthén & Muthén.

Nærde, A., Ogden, T., Janson, H., & Zachrisson, H. D. (2014). Normative development of physical aggression from 8 to 26 months. *Developmental Psychology*, *50*, 1710–1720. doi:10.1037/a0036324

NICHD Early Child Care Research Network. (2004). Trajectories of physical aggression from toddlerhood to middle childhood: Predictors, correlates, and outcomes. *Monographs of the Society for Research in Child Development*, *69*, 1–143.

Olson, I. R., & Marshuetz, C. (2005). Facial attractiveness is appraised in a glance. *Emotion*, *5*, 498–502. doi:10.1037/1528-3542.5.4.498

Olweus, D. (1979). Stability of aggressive reaction patterns in males: A review. *Psychological Bulletin*, *86*, 852–875. doi:10.1037/0033-2909.86.4.852

Olweus, D. (2013). School bullying: Development and some important challenges. *Annual Review of Clinical Psychology*, *9*, 751–780. doi:10.1146/annurev-clinpsy-050212-185516

O'Toole, S. E., Monks, C. P., & Tsermentseli, S. (2017). Executive function and theory of mind as predictors of aggressive and prosocial behavior and peer acceptance in early childhood. *Social Development*, 1–13. doi:10.1111/sode.12231

Otterpohl, N., Imort, S., Lohaus, A., & Heinrichs, N. (2012). Kindliche Regulation von Wut: Effekte familiärer Kontextfaktoren [Anger regulation in children: Effects of contextual factors]. *Kindheit und Entwicklung*, *21*, 47–56. doi:10.1026/0942-5403/a000069

Parker, E. H., Hubbard, J. A., Ramsden, S. R., Relyea, N., Dearing, K. F., Smithmyer, C. M., & Schimmel, K. D. (2001). Children's use and knowledge of display rules for anger following hypothetical vignettes versus following live peer interaction. *Social Development*, *10*, 528–557. doi:10.1111/1467-9507.00179

Patzer, G. L. (1994). Measurement of physical attractiveness: Truth-of-Consensus. *Journal of Esthetic and Restorative Dentistry*, *6*, 185–188. doi:10.1111/j.1708-8240.1994.tb00856.x

Patzer, G. L. (2006). *The power and paradox of physical attractiveness*. Boca Raton, FL, US: Brown Walker Press.

Perrett, D. (2012). *In your face: The new science of human attraction* (1st ed.). Basingstoke, UK: Palgrave Macmillan.

Peterson, C. C., Wellman, H. M., & Slaughter, V. (2012). The mind behind the message: Advancing theory-of-mind scales for typically developing children, and those with deafness, autism, or Asperger syndrome. *Child Development*, *83*, 469–485. doi:10.1111/j.1467-862 4.2011.01728.x

Phillips, L. H., Henry, J. D., Hosie, J. A., & Milne, A. B. (2006). Age, anger regulation and well-being. *Aging & Mental Health*, *10*, 250–256. doi:10.1080/13607860500310385

Pixability. (2014). Beauty on YouTube: How YouTube is radically transforming the beauty industry and what that means for brands. Retrieved from http://www.pixability.com/wp-content/uploads/2014/01/Pixability_Beauty_Study_vf_2.4.2014.pdf

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *The Journal of Applied Psychology*, *88*, 879–903. doi:10.1037/0021-9010.88.5.879

Potegal, M., Stemmler, G., & Spielberger, C. (Eds.). (2010). *International handbook of anger: Constituent and concomitant biological, psychological, and social processes*. New York, NY, US: Springer.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*, 515–526. doi:10.1017/S0140525X00076512

Rauer, W., & Schuck, K. D. (2003). *FEESS 3-4: Fragebogen zur Erfassung emotionaler und sozialer Schulerfahrungen von Grundschulkindern dritter und vierter Klassen* [FEESS 3-4: Questionnaire on social and emotional experiences at school of elementary school children in third and fourth grade]. Göttingen, D: Beltz Test GmbH.

Rauer, W., & Schuck, K. D. (2004). *FEESS 1-2: Fragebogen zur Erfassung emotionaler und sozialer Schulerfahrung von Grundschulkindern erster und zweiter Klassen* [FEESS 1-2: Questionnaire on social and emotional experiences at school of elementary school children in first and second grade]. Göttingen, D: Beltz Test GmbH.

Raval, V. V., Martini, T. S., & Raval, P. H. (2007). 'Would others think it is okay to express my feelings?' Regulation of anger, sadness and physical pain in Gujarati children in India. *Social Development*, *16*, 79–105. doi:10.1111/j.1467-9507.2007.00373.x

Reebye, P. (2005). Aggression during early years: Infancy and preschool. *The Canadian Child and Adolescent Psychiatry Review*, *14*, 16–20.

Reijntjes, A., Kamphuis, J. H., Prinzie, P., Boelen, P. A., van der Schoot, M., & Telch, M. J. (2011). Prospective linkages between peer victimization and externalizing problems in children: A meta-analysis. *Aggressive Behavior*, *37*, 215–222. doi:10.1002/ab.20374

Renouf, A., Brendgen, M., Parent, S., Vitaro, F., David Zelazo, P., Boivin, M.,. . . Séguin, J. R. (2010). Relations between theory of mind and indirect and physical aggression in kindergarten: Evidence of the moderating role of prosocial behaviors. *Social Development*, *19*, 535–555. doi:10.1111/j.1467-9507.2009.00552.x

Rieger, G., Gygax, L., Linsenmeier, J. A. W., Siler-Knogl, A., Moskowitz, D. A., & Bailey, J. M. (2011). Sex typicality and attractiveness in childhood and adulthood: assessing their relationships from videos. *Archives of Sexual Behavior*, *40*, 143–154. doi:10.1007/s10508-009-9512-8

Roberton, T., Daffern, M., & Bucks, R. S. (2012). Emotion regulation and aggression. *Aggression and Violent Behavior*, *17*, 72–82. doi:10.1016/j.avb.2011.09.006

Rohlf, H., & Krahé, B. (2015). Assessing anger regulation in middle childhood: Development and validation of a behavioral observation measure. *Frontiers in Psychology*, *6*, Article 453. doi:10.3389/fpsyg.2015.00453

Röll, J., Koglin, U., & Petermann, F. (2012). Emotion regulation and childhood aggression: Longitudinal associations. *Child Psychiatry and Human Development*, *43*, 909–923. doi:10.1007/s10578-012-0303-4

Rosen, L. H., Underwood, M. K., & Beron, K. J. (2011). Peer victimization as a mediator of the relation between facial attractiveness and internalizing problems. *Merrill-Palmer Quarterly*, *57*, 319–347. doi:10.1353/mpq.2011.0016

Saleem, M., Anderson, C. A., & Barlett, C. P. (2015). Assessing helping and hurting behaviors through the Tangram Help/Hurt Task. *Personality & Social Psychology Bulletin*, *41*, 1345–1362. doi:10.1177/0146167215594348

Salisch, M. von. (2000). *Wenn Kinder sich ärgern: Emotionsregulation in der Entwicklung* [When children get angry: Emotional regulation in the course of development]. Göttingen, D: Hogrefe.

Savolainen, R. (2014). Emotions as motivators for information seeking: A conceptual analysis. *Library & Information Science Research*, *36*, 59–65. doi:10.1016/j.lisr.2013.10.004

Schaie, K. W., & Hofer, S. M. (2001). Longitudinal studies in aging research. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (5th ed., pp. 53–77). San Diego, CA, US: Academic Press.

Schein, S. S., & Langlois, J. H. (2015). Unattractive infant faces elicit negative affect from adults. *Infant Behavior and Development*, *38*, 130–134. doi:10.1016/j.infbeh.2014.12.009

Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review*, *115*, 336–356. doi:10.1037/0033-295X.115.2.336

Sebastian, C. L., Fontaine, N. M. G., Bird, G., Blakemore, S.-J., Brito, S. A. D., McCrory, E. J. P., & Viding, E. (2012). Neural processing associated with cognitive and affective theory of mind in adolescents and adults. *Social Cognitive and Affective Neuroscience*, *7*, 53–63. doi:10.1093/scan/nsr023

Selig, J. P., & Little, T. D. (2012). Autoregressive and cross--lagged panel ynalysis for longitudinal data. In B. P. Laursen, T. D. Little, & N. A. Card (Eds.), *Handbook of developmental research methods* (pp. 265–278). New York, NY, US: Guilford Press.

Serketich, W. J., & Dumas, J. E. (1997). Adults' perceptions of the behavior of competent and dysfunctional children based on the children's physical appearance. *Behavior Modification*, *21*, 457–469. doi:10.1177/01454455970214004

Shamay-Tsoory, S. G., Harari, H., Aharon-Peretz, J., & Levkovitz, Y. (2010). The role of the orbitofrontal cortex in affective theory of mind deficits in criminal offenders with psychopathic tendencies. *Cortex*, *46*, 668–677. doi:10.1016/j.cortex.2009.04.008

Shields, S. A. (2005). The politics of emotion in everyday life: "Appropriate" emotion and claims on identity. *Review of General Psychology*, *9*, 3–15. doi:10.1037/1089-2680.9.1.3

Slater, A., Schulenburg, C. von der, Brown, E., Badenoch, M., Butterworth, G., Parsons, S., & Samuels, C. (1998). Newborn infants prefer attractive faces. *Infant Behavior and Development*, *21*, 345–354. doi:10.1016/S0163-6383(98)90011-X

Slaughter, V. (2015). Theory of mind in infants and young children: A review. *Australian Psychologist*, *50*, 169–172. doi:10.1111/ap.12080

Sparks, D. (2016). Reducing stereotype threat in the science and mathematics classroom: An overview of research, best practices, and intervention strategies. *Currents in Teaching & Learning*, *7*, 4–17.

Stemmler, G. (2010). Somatovisceral activation during anger. In M. Potegal, G. Stemmler, & C. Spielberger (Eds.), *International handbook of anger. Constituent and concomitant biological, psychological, and social processes* (pp. 103–121). New York, NY, US: Springer.

Sutton, J., Smith, P. K., & Swettenham, J. (1999a). Social cognition and bullying: Social inadequacy or skilled manipulation? *British Journal of Developmental Psychology*, *17*, 435–450. doi:10.1348/026151099165384

Sutton, J., Smith, P. K., & Swettenham, J. (1999b). Bullying and 'theory of mind': A critique of the 'social skills deficit' view of anti-social behaviour. *Social Development*, *8*, 117–127. doi:10.1111/1467-9507.00083

Swami, V., & Furnham, A. (2008). *The psychology of physical attraction*. London, UK: Routledge.

Sweeting, H., & West, P. (2010). Being different: Correlates of the experience of teasing and bullying at age 11. *Research Papers in Education*, *16*, 225–246. doi:10.1080/02671520110058679

Tamir, M. (2016). Why do people regulate their emotions? A taxonomy of motives in emotion regulation. *Personality and Social Psychology Review*, *20*, 199–222. doi:10.1177/1088868315586325

Tan, P. Z., Armstrong, L. M., & Cole, P. M. (2013). Relations between temperament and anger regulation over early childhood. *Social Development*, *22*, 755–772. doi:10.1111/j.1467-9507.2012.00674.x

Trampe, D., Quoidbach, J., & Taquet, M. (2015). Emotions in everyday life. *PLOS ONE*, *10*, e0145450. doi:10.1371/journal.pone.0145450

Umberson, D., & Montez, J. K. (2010). Social relationships and health: A flashpoint for health policy. *Journal of Health and Social Behavior*, *51*, 54-66. doi:10.1177/0022146510383501

Underwood, M. K. (1997). Top ten pressing questions about the development of emotion regulation. *Motivation and Emotion*, *21*, 127–146. doi:10.1023/A:1024482516226

Valois, R. F., MacDonald, J. M., Bretous, L., Fischer, M. A., & Drane, J. W. (2002). Risk factors and behaviors associated with adolescent violence and aggression. *American Journal of Health Behavior*, *26*, 454–464. doi:10.5993/AJHB.26.6.6

Vigil-Colet, A., Ruiz-Pamies, M., Anguiano-Carrasco, C., & Lorenzo-Seva, U. (2012). The impact of social desirability on psychometric measures of aggression. *Psicothema*, *24*, 310–315.

Waters, S. F., & Thompson, R. A. (2014). Children's perceptions of the effectiveness of strategies for regulating anger and sadness. *International Journal of Behavioral Development*, *38*, 174–181. doi:10.1177/0165025413515410

Waters, S. F., & Thompson, R. A. (2016). Children's perceptions of emotion regulation strategy effectiveness: Links with attachment security. *Attachment & Human Development*, *18*, 354–372. doi:10.1080/14616734.2016.1170051

Wheeler, L., & Kim, Y. (1997). What is beautiful is culturally good: The physical attractiveness stereotype has different content in collectivistic cultures. *Personality and Social Psychology Bulletin*, *23*, 795–800. doi:10.1177/0146167297238001

Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention. Methodological advances from alcohol and substance abuse research* (pp. 281–324). Washington, DC, US: American Psychological Association. doi:10.1037/10222-009

Wilson, S. J., & Lipsey, M. W. (2007). School-based interventions for aggressive and disruptive behavior: Update of a meta-analysis. *American Journal of Preventive Medicine*, *33*, S130-S143. doi:10.1016/j.amepre.2007.04.011

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*, 103–128. doi:10.1016/0010-0277(83)90004-5

Appendix

Appendix A: Cross-Lagged Panel Analyses of Anger Regulation and Aggression

Appendix B: Self-Reinforcing Cycles of Aggression via Theory of Mind

Appendix C: Central Instruments Used in the Empirical Papers

Appendix A: Cross-Lagged Panel Analyses of Anger Regulation and Aggression

Paper 1 had a methodological focus in examining the structural equivalence and validity of the two behavioral observation assessments of maladaptive anger regulation. The examination of the predictive value of anger regulation for aggressive behavior was only marginally done by checking bivariate correlations between maladaptive anger regulation at T1 and aggression measures at T3. Specifically, at T3 there were a teacher-report, a child-report, and a behavioral measure of aggressive behavior. Here, cross-lagged panel analyses are presented that are assumed to go some way toward a causal interpretation of the pathways (Marmor & Montemayor, 1977). These analyses are done only for the teacher-report of aggressive behavior as only this measure was used at both T1 and T3. In Paper 1, maladaptive anger regulation was correlated concurrently with teacher-reported aggression at T1 ($r = .20$, $p < .001$) and T3 ($r = .22$, $p < .05$), but not predictively from anger regulation at T1 to aggression at T3 ($r = .06$, $p = .35$).

The first cross-lagged panel model considered teacher-reported aggressive behavior as one scale of overall aggression (*overall model*). The second cross-lagged panel model distinguished between the relational and the physical form of aggression (*differentiated model*). Both models were controlled for age and sex. In both models, latent factors of maladaptive anger regulation at T1 and T3 were modelled under the partial metric invariance assumption (strategy solution orientation was freed; for a more detailed description, see Paper 1). In both models, latent factors of aggressive behavior and their subforms, respectively, were modelled under the strict metric invariance assumption. Statistical analyses were based on the multiply imputed data sets using the WLSMV estimator (see Paper 1).

The overall model showed an acceptable fit with the data ($\chi^2[241] = 461.43$, $p < .001$; RMSEA = .04; WRMR = 1.14; CFI = .93; TLI = .92; see Figure AP-1). Contrary to expectations, the path from maladaptive anger regulation at T1 to overall aggression at T3 was negative and nonsignificant ($\beta = -.04$, $p = .54$). The inverse path from aggression at T1 to maladaptive anger regulation at T3 was also nonsignificant ($\beta = -.01$, $p = .91$). Stabilities were of moderate size for both constructs ($\beta s \geq .47$, $ps < .001$). The cross-sectional correlation between the constructs was only significant at T1 ($r = .18$, $p < .01$), but not at T3 ($r = .18$, $p = .08$).

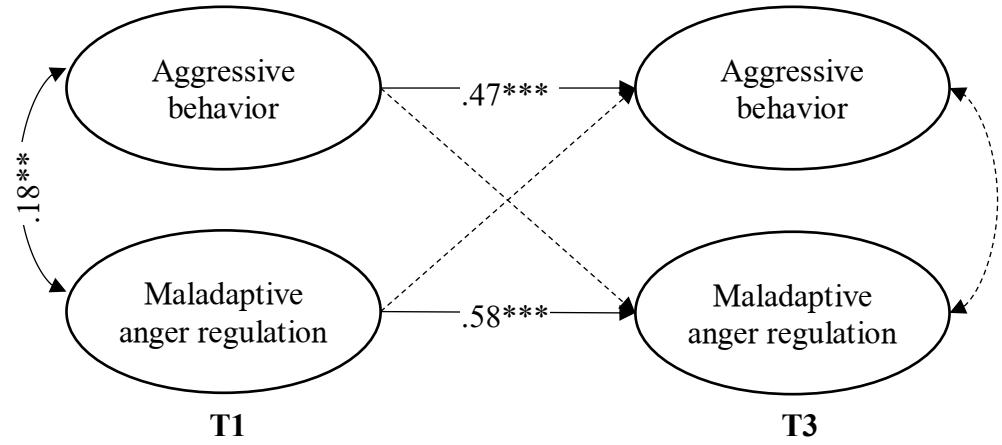

Figure AP-1. Cross-lagged panel model for maladaptive anger regulation and overall aggression (overall model)

Notes. Standardized coefficients are displayed; dashed lines display nonsignificant paths; all variables are controlled for age and sex.

The differentiated model also showed an acceptable fit with the data ($\chi^2[236] = 445.06$, $p < .001$; RMSEA = .04; WRMR = 1.10; CFI = .94; TLI = .93; see Figure AP-2). Consistent with the overall model, the paths from maladaptive anger regulation at T1 to both forms of aggression at T3 were negative and nonsignificant (βs = -.03, *p*s ≥ .59). Again, there were no significant inverse paths from aggression to anger regulation (βs ≤ .09, *p*s ≥ .31). However, physical aggression at T1 was predictive of relational aggression at T3 (β = .20, $p < .01$). Stabilities were of moderate size for maladaptive anger regulation and physical aggression (βs ≥ .43, *p*s < .001), but low for relational aggression (β = .27, $p < .001$). The constructs' cross-sectional correlation was significantly positive (*r*s ≥ .15, *p*s < .05), with an exception of maladaptive anger regulation and relational aggression at T3 ($r = .12$, $p = .23$).

As a final step, potential sex differences in the path from maladaptive anger regulation at T1 to aggression at T3 in both presented models were tested. Traditional model comparisons by comparing a constrained versus an unconstrained model based on chi-square differences cannot be used when using multiply imputed data sets (Asparouhov & Muthén, 2010b). Thus, sex differences in the path of interest (i.e., anger regulation T1 to aggression T3) were examined using Wald tests. In neither model, there were significant differences (overall model: $W = 0.71$, $p = .40$; differentiated model: $W_{physical} = 0.16$, $p = .69$; $W_{relational} = 1.04$, $p = .31$). Thus, the non-existence of a prospective path from anger regulation to aggression holds for both sexes.

In conclusion, the presented analyses (overall and differentiated model) revealed that anger regulation and (teacher-reported) aggressive behavior are correlated with one another.

However, the analyses do not support the notion that anger regulation has a predictive value for the development of aggression in childhood.

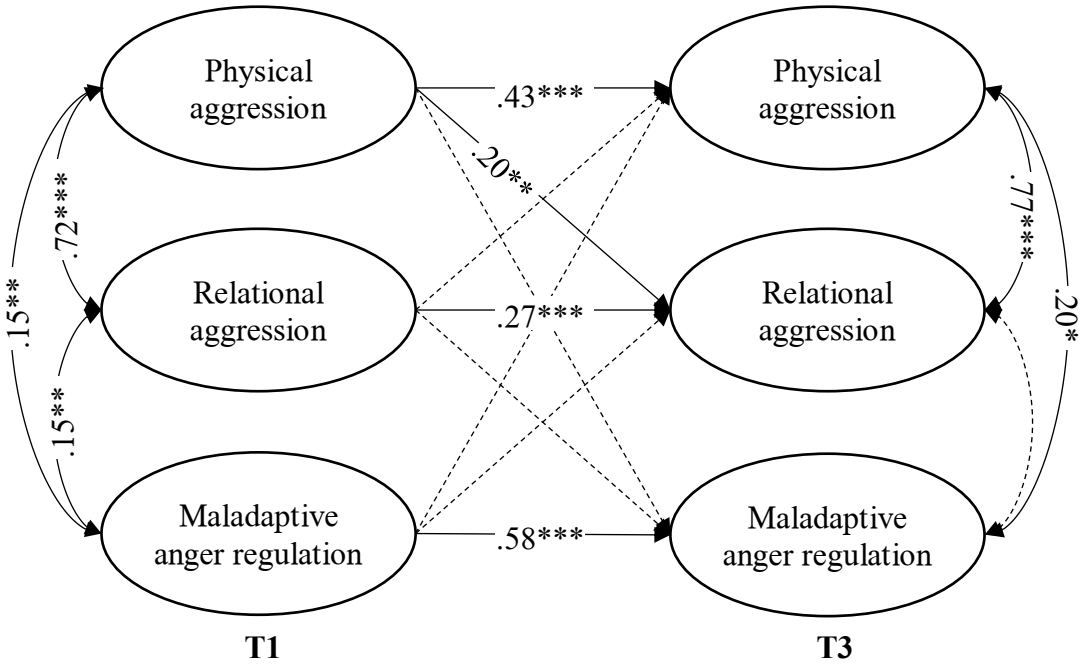

Figure AP-2. Cross-lagged panel model for maladaptive anger regulation and aggression forms (differentiated model)

Notes. Standardized coefficients are displayed; dashed lines display nonsignificant paths; all variables are controlled for age and sex.

Appendix B: Self-Reinforcing Cycles of Aggression via Theory of Mind

In Paper 2, reciprocal effects between aggressive behavior and theory of mind over time were found. Specifically, relational aggression at T1 was predictive of theory of mind at T2 ($\beta = -.12$, $p < .05$), and theory of mind at T2 was predictive of both relational ($\beta = -.11$, $p < .05$) and physical aggression ($\beta = -.12$, $p < .01$) at T3. The combination of these paths suggests the existence of self-reinforcing cycles of aggression through theory of mind development. Here, this notion is tested by computing indirect effects using R (Version 3.3.1) and following a Monte Carlo procedure with 20,000 repetitions (Hayes & Scharkow, 2013).

The indirect path from relational aggression at T1 to relational aggression at T3 through theory of mind at T2 was significant ($\beta = .012$, 95% CI [.001, .032]) as well as the path from relational aggression at T1 to physical aggression at T3 through theory of mind at T2 ($\beta = .013$, 95% CI [.001, .033]). Thus, there is evidence for self-reinforcing cycles of aggression in this data. Children, who were more relationally aggressive at T1, were found to be less skilled in theory of mind nine months later at T2, which in turn predicted more aggressive behavior (relationally and physically) further two years later at T3.

Appendix C: Central Instruments Used in the Empirical Papers

Note: Text in italics displays verbatim wording.

Aggression via teacher-report (T1, T2, T3)

Format:

Questionnaire (adapted from the Children´s Social Behavior Scale – Teacher Form, CSBS-T, Crick, 1996)

Instruction:

Im folgenden Fragebogen geht es um verschiedene Formen von Aggression und Gründe für dieses aggressive Verhalten. Möglicherweise können Sie nicht immer genau sagen, wie oft oder aus welchen Gründen ein Kind etwas getan hat. Versuchen Sie dann bitte eine Antwort zu geben, von der Sie denken, dass sie so nah wie möglich an die wirkliche Situation herankommt.

Items:

Wie oft hat das Kind in den letzten sechs Monaten …

(1) andere Kinder geschlagen, geschubst oder gestoßen?[a]

(2) damit gedroht, andere Kinder zu schlagen oder zu verprügeln?[a]

(3) körperliche Auseinandersetzungen mit anderen Kindern initiiert oder ist es in solche hineingeraten?[a]

(4) Gerüchte über andere Kinder verbreitet oder über diese getratscht?[b]

(5) versucht, ein anderes Kind von Gruppenaktivitäten auszuschließen?[b]

(6) einem anderem Kind damit gedroht, nicht mehr sein Freund zu sein?[b]

[a] items for physical aggression, [b] items for relational aggression

Response scale:

1 – nie

2 – einmal im Monat oder seltener

3 – mehrmals pro Monat

4 – mehrmals pro Woche

5 – täglich

Aggression via children's self-report (T2, T3)

Format:

Questionnaire (adapted from the Children´s Social Behavior Scale – Teacher Form, CSBS-T, Crick, 1996)

Instruction:

Bei diesen Fragen geht es um dein Verhalten gegenüber Mitschülern und um die Gründe für dieses Verhalten.

Items:

Wie oft hast du in den letzten sechs Monaten …

(1) andere Kinder geschlagen, geschubst oder gestoßen?[a]

(2) einem anderen Kind damit gedroht, es zu schlagen oder zu verprügeln?[a]

(3) eine Prügelei angefangen oder bei einer Prügelei mitgemacht?[a]

(4) über andere Kinder gelästert (schlecht über sie geredet)?[b]

(5) ein anderes Kind nicht mitspielen lassen?[b]

(6) einem anderem Kind damit gedroht, dass du nicht mehr sein Freund sein wirst?[b]

[a] items for physical aggression, [b] items for relational aggression

Response scale:

1 – nie oder fast nie

2 – einmal im Monat

3 – einmal in der Woche

4 – fast jeden Tag

Aggression via behavioral measure (T3)

Format:

One-item measure (self-constructed, but based on the Tangram Hurt/Help Task, Saleem et al., 2015)

Instruction:

[directly after anger-eliciting task]

Pass auf, du kannst jetzt entscheiden, wie viele Würfel eines der Kinder an der Schule, an die wir als nächstes gehen, für dieselbe Aufgabe bekommen soll. Du hast den Turm mit 7 Würfeln gestapelt, wir haben höchstens 12 Würfel zur Verfügung.

Item:

Wie viele Würfel soll deiner Meinung nach ein anderes Kind bekommen, um einen solchen Turm zu stapeln?

Response scale:

2-12 Würfel

transformed to

0 – no increase of the difficulty level (i.e., 7 dice or fewer)

1 – increase by one die (i.e., 8 dice)

2 – increase by two dice (i.e., 9 dice)

3 – increase by three dice (i.e., 10 dice)

4 – increase by four dice (i.e., 11 dice)

5 – highest increase of the difficulty level (i.e., maximum number of 12 dice)

Anger regulation (T1, T3)

Format:

Behavioral observation (coding of strategy use during an anger-eliciting task)

Instruction T1:

Jetzt habe ich eine kleine Aufgabe für dich. Du sollst aus diesen Bauklötzen hier den Turm auf diesem Foto nachbauen. Du hast drei Minuten Zeit, an dieser Sanduhr kannst du sehen, wieviel Zeit dir noch bleibt. Wenn du es in dieser Zeit schaffst den Turm aufzubauen, darfst du dir eines dieser Geschenke aussuchen. Der Turm muss einen Moment ohne Festhalten stehenbleiben. Außerdem gibt es noch ein großes Geschenk, das das Kind bekommt, das von allen Kindern das schnellste war. Ok, dann geht es los, wenn ich die Sanduhr gleich umdrehe. Einen Moment... Und Los!

Instruction T3:

Jetzt habe ich eine kleine Aufgabe für dich. Du sollst diese Würfel hier zu einem Turm stapeln. Das kann zum Beispiel so aussehen, wie auf diesem Foto hier, der Turm kann aber auch ganz anders aussehen. Du hast dafür zwei Minuten und 40 Sekunden Zeit. Anhand dieser Uhr hier kannst du sehen, wieviel Zeit dir noch bleibt. Wenn du es in dieser Zeit schaffst den Turm aufzubauen, darfst du dir eines dieser Geschenke aussuchen. Der Turm muss einen Moment ohne Festhalten stehenbleiben. Außerdem gibt es noch ein großes Geschenk, das das Kind bekommt, das von allen Kindern das schnellste war. Ok, ist soweit alles in Ordnung? ... Dann geht es gleich los, sobald ich die Uhr starte. Einen Moment ... Und Los!

Response coding:

Visual focus on frustrating stimuli (maladaptive strategy)

- Looking at the timer[a]
- Looking at the gifts[a]

Verbal focus on frustrating stimuli (maladaptive strategy)

- Talking negatively about the time[a]
- Talking negatively about the gifts[a]
- Talking negatively about the task[a]
- Talking negatively about the self[a]

Venting the anger (maladaptive strategy)

- Verbal expression of anger[a]
- Anger expression in mimic and gesture[a]
- Handling the material roughly[a]

Resignation (maladaptive strategy)

- Refusing to continue for at least 3 sec (0 vs. 1)[b]

Solution orientated behavior T1 (adaptive strategy)

- Testing an alternative strategy[a]
- Balancing out the tower (in sec)[c]
- Working in a focused way (1-4)[d]

Solution orientated behavior T3 (adaptive strategy)

- Testing an alternative strategy (0-5)[e]
- Balancing out the tower[a]
- Working in a focused way (1-4)[d]
- Rearranging dice[a]
- Following the example picture (0 vs. 1)[b]

[a] number of occurrences (i.e., event-coding), [b] dichotomous (no vs. yes), [c] measure of duration (i.e., time-coding), [d] rating by coder, [e] number of first occurrences

Theory of Mind (T1, T2)

Format:

Computer-assisted task (responses to cartoon-stories, adapted from Sebastian et al., 2012)

Instruction:

Wir werden uns gleich zusammen eine Bildergeschichte auf dem Computer anschauen. Sie besteht aus mehreren Bildern, die nacheinander auf dem Bildschirm erscheinen. Nachdem dir drei Bilder der Geschichte gezeigt wurden, siehst du ein Fragezeichen. Danach erscheinen zwei Endungen, die möglich sind – eine richtige und eine falsche. Deine Aufgabe ist es, das richtige Ende der Geschichte so schnell wie möglich auszuwählen. Um das linke Bild zu wählen, drückst du die linke blaue Taste [X-key]*, um das rechte zu wählen, die rechte blaue Taste* [M-key]*. Wir üben das jetzt erst einmal etwas langsamer und ohne die Tasten zu drücken.*

[two exercise trials]

So, dann kann es jetzt losgehen. Bitte denke daran, das richtige Ende der Geschichte so schnell wie möglich zu wählen. Bist du bereit? Ok.

Trials:

affective theory of mind	cognitive theory of mind
(1) cat	(1) umbrella
(2) boat	(2) butterfly
(3) hammer	(3) ladder
(4) football	(4) door
(5) storm	(5) bakery
(6) kite	(6) bag

Response coding:

0 – incorrect response

1 – correct response

Physical attractiveness (T1, T3)

Format:

Rating (assessed by 24 [T1] and 23 [T3] independent adult raters based on facial photo of the child)

Instruction:

Im Folgenden werden Ihnen Fotos von Kindern im Alter von 6 bis 10 Jahren [T3: *von 9 bis 13 Jahren*] *präsentiert. Ihre Aufgabe wird es sein, diese Kinder hinsichtlich ihrer physischen Attraktivität zu beurteilen. Im umgangssprachlichen Gebrauch spricht man bei physischer Attraktivität auch oft von „gutem Aussehen" oder „hübsch sein".*

Kinder nach Ihrer physischen Attraktivität zu beurteilen wirkt für viele Menschen ein wenig befremdlich. Dennoch unterscheiden sich auch Kinder in ihrem Aussehen, und damit in dem Grad ihrer Attraktivität. Bitte schätzen Sie daher jedes Kind hinsichtlich seiner physischen Attraktivität auf einer Skala von 1 gar nicht bis 7 sehr ein.

Da es sich bei den Fotos um Screenshots aus Videoaufnahmen handelt, kann der Winkel, die Bildqualität und die Beleuchtung der Fotos variieren. Auch der Gesichtsausdruck und der Blick der Kinder unterscheiden sich zwischen den Fotos teilweise leicht. Bitte lassen Sie sich von diesen Uneinheitlichkeiten nicht irritieren, und versuchen Sie Ihr Urteil unabhängig von diesen Aspekten abzugeben.

Sollten Sie Ihr Urteil noch einmal ändern wollen, haben Sie die Möglichkeit mit dem Klick auf den Button „Zurück" zu dem vorigen Foto zurückzukehren. Sollten Sie das Kind kennen (z.B. aus einer vergangenen Testung der PIER-Studie als Testleiter/in), klicken Sie bitte "Ich kenne das Kind" an. Sollte Ihnen kein Foto angezeigt werden, melden Sie sich bitte bei dem Versuchsleiter. Sollte das Bild eine zu niedrige Qualität aufweisen, sodass Sie das Kind nicht erkennen können, klicken Sie bitte auf "Kind nicht erkennbar".

Vielen Dank für Ihre Unterstützung!

Item:

[for each facial photo]

Wie attraktiv/hübsch ist dieses Kind?

Response scale:

1 – gar nicht
2 | 3 | 4 | 5 | 6
7 – sehr

Peer problems via teacher-report (T1, T2, T3)

Format:

Questionnaire (items from the subscale Peer Problems of the Strengths and Difficulties Questionnaire, SDQ, Goodman, 1997, combined with one self-constructed item)

Instruction:

Bitte geben Sie an, inwieweit die folgenden Aussagen in den letzten sechs Monaten auf das Kind zutrafen.

Items:

Das Kind …

(1) ist im Allgemeinen bei anderen Kindern beliebt.[a]

(2) wird von anderen gehänselt oder schikaniert.[a]

(3) wird oft ausgeschlossen, wenn seine Klassenkameraden in den Pausen etwas zusammen machen.[b]

[a] items of the SDQ, [b] self-constructed item

Response scale:

1 – nicht zutreffend

2 – teilweise zutreffend

3 – eindeutig zutreffend

Peer problems via children's self-report (T1, T2, T3)

Format:

Questionnaire (items from the subscale Social Integration of the Questionnaire on Social and Emotional Experiences at School of Elementary School Children, FEESS, Rauer & Schuck, 2003, 2004; at T1 and T2 combined with items from the subscale Peer Acceptance of the German version of the Harter-Scales, Asendorpf & van Aken, 1993)

Instruction:

Die folgenden Sätze drücken aus, was Schülerinnen und Schüler denken und fühlen.

Items:

At all measurement points

(1) Meine Mitschüler sind nett zu mir.[a]

(2) Meine Mitschüler trösten mich, wenn ich traurig bin.[a]

(3) Die anderen lachen mich häufig aus.[a]

(4) Ich darf beim Spielen auf dem Schulhof mitmachen.[a]

(5) Die anderen suchen Streit mit mir.[a]

Only at T1 and T2

(6) Andere Kinder möchten gerne neben mir sitzen .[b]

(7) Ich werde oft gefragt, ob ich mitspielen möchte.[b]

(8) Die anderen Kinder mögen mich gerne.[b]

Only at T3

(6) Ich habe wenig Freunde in der Klasse.[a]

(7) Ich komme mit den anderen Kindern in meiner Klasse gut aus.[a]

[a] items of the FEESS, [b] items of the Harter Scales

Response scale:

At T1 and T2	At T3
1 – nein	*1 – nein*
2 – ja	*2 – eher nein*
	3 – eher ja
	4 – ja

Peer problems via parent-report (T1, T2, T3)

Format:

Questionnaire (subscale Peer Problems of the Strengths and Difficulties Questionnaire, SDQ, Goodman, 1997)

Instruction:

Bitte geben Sie an, wie sehr die folgenden Punkte auf Ihr Kind zutreffen. Bitte berücksichtigen Sie bei der Antwort das Verhalten Ihres Kindes in den letzten sechs Monaten.

Items:

(1) Einzelgänger; spielt meist alleine

(2) Hat wenigstens einen guten Freund oder eine gute Freundin

(3) Im Allgemeinen bei anderen Kindern beliebt

(4) Wird von anderen gehänselt oder schikaniert

(5) Kommt besser mit Erwachsenen aus als mit anderen Kindern

Response scale:

1 – nicht zutreffend

2 – teilweise zutreffend

3 – eindeutig zutreffend

Erklärung

Hiermit erkläre ich, dass die Dissertation „Intrapersonal risk factors of aggressive behavior in childhood: A longitudinal perspective“ selbstständig und ohne Hilfe Dritter verfasst wurde und bei der Abfassung alle Regelungen guter wissenschaftlicher Standards eingehalten wurden. Ich erkläre, dass die Dissertation in der gegenwärtigen Fassung keiner anderen Hochschule zur Begutachtung vorgelegen hat oder vorliegt. Ich erkläre, an keiner anderen Hochschule ein Promotionsverfahren eröffnet zu haben.

Potsdam, June 2017

_____

Fabian Kirsch