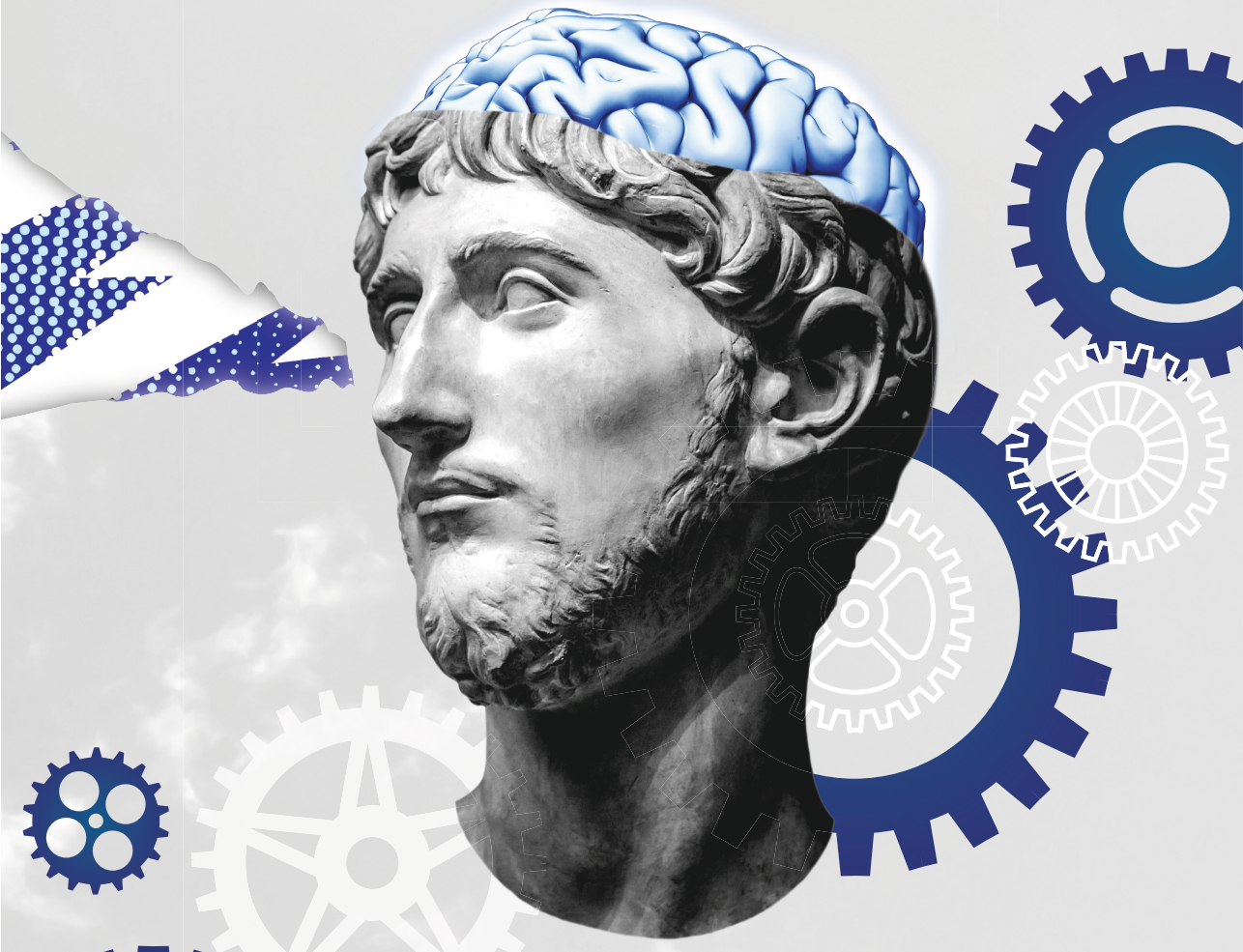
# Cognitive Bias Modification for Aggression-Related Biases of Attention and Interpretation



Nouran AlMoghrabi

# Cognitive Bias Modification for Aggression-Related Biases of Attention and Interpretation

Nouran AlMoghrabi

Cover design by Daniëlle Balk, www.persoonlijkproefschrift.nl
Layout by Ridderprint BV, www.ridderprint.nl
Printed by Ridderprint BV, www.ridderprint.nl

ISBN: 978-94-6375-770-6

The research presented in this dissertation was financially supported by Princess Nourah bint Abdulrahman University

# Cognitive Bias Modification for Aggression-Related Biases of Attention and Interpretation

Cognitieve bias modificatie voor agressie-gerelateerde vertekeningen in aandacht en interpretatie

**Proefschrift**

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de
rector magnificus

Prof.dr. R.C.M.E. Engels

en volgens besluit van het College voor Promoties.
De openbare verdediging zal plaatsvinden op
donderdag 5 maart 2020 om 13:30

door

Nouran AlMoghrabi
geboren te Dammam, Saoedi-Arabië

Erasmus University Rotterdam

**Promotiecommissie:**

| | |
|---|---|
| **Promotor:** | Prof. dr. I.H.A. Franken |
| **Overige leden:** | Prof. dr. B. Orobio de Castro<br>Prof. dr. E.G.C. Rassin<br>Dr. E. Salemink |
| **Copromotoren:** | Dr. J. Huijding<br>Dr. B. Mayer |

# CONTENTS

# CHAPTER 1

## General Introduction

Our daily lives are full of clumsy social interactions: we bump into each other in crowded hallways, we spill drinks on clothes and belongings, we close doors in faces, and we make clumsy remarks. These kinds of incidents happen all the time, and navigating these episodes in a socially acceptable manner is not always easy. While such incidents are often unintentional, and reactive aggression would be inappropriate and counterproductive, we sometimes react with aggression, and some people do so more than others. Aggression in our societies is a serious and growing problem (Krug, Mercy, Dahlberg, & Zwi, 2002), imposing negative emotional, physical, and economic consequences on aggressive individuals, their victims, their families, and the larger society (de Castro, 2004; Krug et al., 2002). Additionally, aggressive individuals are at risk for various negative outcomes, such as academic failure and dropping out of school (Hymel, Comfort, Schonert-Reichl, & McDougall, 1996), criminal behavior (Swogger, Walsh, Christie, Priddy, & Conner, 2015), social difficulties (Dodge & Coie, 1987), relationship problems (Curtis, Epstein, & Wheeler, 2017), substance abuse (Skara et al., 2008), low self-esteem (Ialongo, Vaden-Kiernan, & Kellam, 1998), and even suicidal attempts (Dumais et al., 2005). In the treatment of aggression, cognitive behavioral therapy (CBT) interventions are commonly used approaches. Despite the fact that CBT is a well-established treatment, its efficacy in treating aggression remains inconsistent among both nonclinical and clinical populations (Lee & DiGiuseppe, 2018). The lack of an effective treatment for aggression calls for a better understanding of the processes underlying aggression in order to improve and develop prevention and intervention programs for individuals with aggressive behavior problems.

A promising line of research has emphasized the role of cognitive biases as a cognitive precursor for maladaptive social behaviors, including trait anger and aggression (Anderson & Bushman, 2002; Crick & Dodge, 1994). Cognitive biases occur when the way information from the internal and external environment is processed leads to systematically distorted representations of the situation compared to objective reality (Haselton, Nettle, & Murray 2015). Depending on situational demands, such biases can be adaptive or maladaptive. In the context of aggression, it has been proposed that biased attention for maladaptive social cues and a tendency to interpret such cues as hostile will lead to hostile representations of social situations and increase the chance of aggressive behavior (e.g., Crick & Dodge, 1994; de Castro, Veerman, Koops, Joop, & Monshouwer, 2002; Wilkowski & Robinson, 2008).

Such findings led to the development of computerized cognitive bias modification (CBM) techniques to modify aggression-related attention and interpretation biases. Although the results of the first studies on the effects of CBM paradigms targeting interpretations (CBM-I) on aggression were promising (e.g., Hawkins & Cougle, 2013; Vassilopoulos, Brouzos, & Andreou, 2015), few studies have examined the effects of CBM on aggression, and all have focused solely on CBM-I. To date, there have not been any studies on the effectiveness of cognitive bias modification paradigms targeting attention (CBM-A). Despite the advances in understanding the role of cognitive biases in aggression, applying this knowledge in (preventive) intervention research targeting aggression is still at its formative stage, and more research regarding the efficacy of these training procedures on both bias and aggression is needed before implementing CBM procedures in therapeutic contexts.

The general focus of the current dissertation is to examine whether a novel CBM procedure using pictorial stimuli can be used to change maladaptive information processing in the context of aggression. In particular, we will focus on changing attentional bias and interpretation bias, and we will explore how these two biases interact. Most importantly, we want to examine the effects of the altered aggression-related cognitive biases on aggressive behavior using self-report and behavioral measures. We aim to establish whether this novel CBM paradigm for aggression is feasible and whether it should target attention, interpretation, or both.

**Aggression and Social Information Processing**

Human aggression can be defined as an intentional behavioral act that is carried out to hurt, harm, or injure another individual (Anderson & Bushman, 2002). Crick and Dodge's model of social information processing (SIP) provides a significant understanding of the development and maintenance of aggression (Crick & Dodge, 1994). Specifically, the model attempts to explain the cognitive process an individual goes through before enacting a behavioral response.

The SIP model proposes that in social situations the most relevant of the diverse social cues are identified and encoded (step 1) and are subsequently used to construct an interpretation of the situation (step 2). After interpreting the situation based on these social cues, the individual formulates goals or outcomes for the situation (step 3). These goals activate familiar responses: responses that are typical for that individual in similar situations (step 4). Those familiar responses are typically stored in long-term memory, or if the situation is new, then they will form new

responses that are most suitable for the situation. After generating multiple responses, those responses are evaluated and the most favorable response is selected (step 5). Finally, the selected response is enacted behaviorally (step 6). These information processes of the SIP model are considered online processes that are related to the processing of the presented social cues, leading to behavioral enactment of those cues. However, the model posited that any step of the online processes may be influenced or guided by offline processes (e.g., social schemas and social knowledge) that an individual has developed from past experiences and events that might serve as a link to individual differences in online processing.

Following this model, maladaptive behaviors, including aggression, may arise from biases during any of the steps of processing social cues, and numerous studies have indeed confirmed that there is a relation between biases in these processes and aggressive behavior (Anderson & Bushman, 2002; Crick & Dodge, 1994). Aggression is multidimensional, and based on the underlying motives for the aggressive act, it can be divided into two subtypes: proactive and reactive (Dodge, 1991). Proactive aggression is a planned, non-provoked behavior, wherein an individual uses aggression to meet a certain goal with the intention to harm another individual. Reactive aggression, on the other hand, is an impulsive angry reaction to a provocation or perceived threat (Poulin & Boivin, 2000). It has been suggested that these different subtypes of aggression are associated with deficits in distinct SIP steps. Researchers propose that biases in encoding and interpreting social cues (step 1 and 2) relate more to reactive aggression (Dodge, 2006). On the other hand, proactive aggression relates more to later stages of the SIP: formulating instrumental goals (Crick & Dodge, 1996), generating alternative responses (Brugman et al., 2015), and evaluating and selecting a specific response to be carried out (Crick & Dodge, 1996).

Aggression studies have focused extensively on the early steps of the SIP model (encoding and interpretation of cues), as these steps elucidate the role of social cues in social situations. The social cues that individuals attend to and the way they disambiguate a situation indicates how they will respond in a social situation. For example, imagine a scenario in which a colleague does not wave back at you as you pass him in the hallway. Encoding not waving back and interpreting the colleague's intention as deliberately ignoring you would lead to a different response than encoding that he was not looking in your direction and interpreting that he was so caught up in his own thoughts that he failed to wave back. Thus, when an individual encodes and interprets another's

intention as hostile, this perception of hostility could justify an "aggressive response" (Dodge & Coie, 1987). Therefore, it has been suggested that the way that aggressive individuals encode (Wilkowski & Robinson, 2008) and interpret a social situation might play a significant role in the etiology and maintenance of aggression (de Castro et al., 2002). Given the significance of the early stages of the SIP model on aggression, the present dissertation is focused on the encoding and interpretation of social cues by further examining the effect of manipulating these cognitive processes on reactive aggression.

**Interpretation Bias in Aggression and Its Modification (CBM-I)**

Although biases in interpretation are the second step of the SIP model, it is this step that has most often been the topic of empirical study. Aggression studies have mostly examined interpretation bias regarding other people's intentions in social situations, often referred to as hostile attribution bias or hostile intent attribution. A hostile attribution bias refers to the tendency to interpret the intentions of others in social situations as hostile, and this tendency is present even if the social situation is ambiguous (Dodge, 1980). A meta-analytic review of these studies confirmed that hostile intent attributions play an important role in the development and maintenance of aggressive behavior (de Castro et al., 2002). When an individual interprets the intentions of others as hostile, this perception of hostility would increase the likelihood of an aggressive response. Furthermore, when an individual acts aggressively toward others, this in turn pushes others to respond aggressively, thus validating the aggressive individual's initial hostile perception of the situation (Crick & Dodge, 1996).

Many studies have examined the relations between hostile intent attributions and behavior problems, including aggression. In a typical experimental design, hostile attribution of intent is assessed by presenting the participant with a number of scenarios of social situations with a hypothetical negative outcome. These scenarios could be presented in written stories or vignettes (e.g., Crick & Dodge, 1996), short video clips (e.g., Dodge & Coie, 1987), or drawn pictures (e.g., Waas, 1988). After the presentation of each scenario, the participants are asked why the other person might have acted the way that he or she did, and they are presented with two response options. Usually one of those responses attributes hostile intent to the other person (i.e., the incident happened on purpose), and the other response attributes prosocial intention to the other person (i.e., the incident happened by accident). Studies using this assessment method typically

found that when aggressive participants are asked to attribute the intention of another's action, they are more likely to interpret the peer's intention as hostile compared to nonaggressive participants (e.g., Crick & Dodge, 1996; Dodge & Coie, 1987). Thus, if interpretation biases are important in the development and maintenance of aggression, then one would expect that a change in interpretation bias would be related to a change in aggression.

Interestingly, a number of studies showed that CBM paradigms can reduce hostile interpretation biases and associated aggressive behavior, thus indicating that CBM-I may find future application in the clinical domain (e.g., Hawkins & Cougle, 2013; Penton-Voak et al., 2013; Vassilopoulos et al., 2014).

CBM-I techniques were initially introduced by Mathews and Mackintosh (2000) and were designed to induce either negative or positive interpretations to reduce symptoms displayed by anxious individuals. These CBM-I paradigms typically modify interpretation bias by repeatedly exposing the participant to ambiguously threatening written vignettes. Depending on the training condition, participants are reinforced for correctly answering questions related to those vignettes either in a negative or a benign way. Similarly, in aggression studies, CBM-I training paradigms repeatedly exposed the participant to ambiguous written vignettes that typically described a social interaction in which something unfortunate happened (i.e., a negative outcome), but, most importantly, the vignettes described an interaction in which the intent of the interacting person is not clear. Each vignette was followed by two or more interpretations. One interpretation or attribution involved a hostile disambiguation of the situation, and the other interpretation involved prosocial or benign disambiguation of the situation. Thus, in aggression literature, this type of training is referred to as either CBM-I or attribution bias modification training since the main focus of this training is giving meaning to the intentions of others (de Castro et al., 2002). Vignette studies have shown that this training method can be successful in increasing prosocial interpretations and decreasing aggression (e.g., Vassilopoulos et al., 2015), as well as increasing hostile interpretations and increasing anger (e.g., Hawkins & Cougle, 2013). However, an issue of concern would be that written vignettes do not fully represent day-to-day interpersonal situations because of the limited amount of contextual information available to the participant. For instance, nonverbal cues, such as facial expressions, contain important situational information regarding the intentions of others (Cadesky, Mota, & Schachar, 2000). Because the vignettes preclude the

possible role of important information, such as facial expressions, in interpreting the intentions of others, they are unlikely to elicit the same hostile attributions as real-life social interactions. Thus, in the current dissertation we wanted to examine the possibility of modifying hostile attribution biases using visual stimuli instead of written vignettes by using images that better reflect what an average person might encounter in their day-to-day life. Each image depicts a social situation in which one character harms another while the intent (intentional or unintentional) of the harm-doer is ambiguous.

The pilot study described in **Chapter 2** examines a novel CBM-I procedure using pictorial stimuli. Male and female university students were trained to interpret ambiguous social situations either in a prosocial or hostile way. Effects on interpretation bias, aggression (self-reported and behavioral measure), anger, and mood were assessed. We expected that training individuals to interpret ambiguous situations in a prosocial way would lead to an increase in prosocial interpretations and a reduction in aggressive behavior whereas training them to interpret such situations as hostile would increase hostile interpretations and aggressive behavior.

Along with adding nonverbal stimuli to the CBM-I training, it is important to experimentally explore the role of these facial expressions in modifying interpretations of intent. It may be the case that in real-life situations hostile interpretation of intent may arise from or occur simultaneously with hostile interpretation of facial expressions and that both biases function as a driving force for aggressive responses. Aggression studies that made use of pictorial stimuli of isolated faces suggested that aggression is associated with interpreting ambiguous facial expressions as hostile (e.g., Schönenberg & Jusyte, 2014; Smeijers, Rinck, Bulten, Van den Heuvel, & Verkes, 2017). Thus far, none of the previous work attempted to integrate both interpretation of intent and interpretation of facial expressions in the training or assessment process. The only study we are aware of is Hiemstra et al. (2018), in which hostile attribution bias was measured after a CBM training that aimed to reduce hostile interpretation of facial expressions. Although the training resulted in changes in the interpretation of facial expressions, those changes did not generalize to changes in interpretations of hostile intent. Thus, further research is needed, as it may be the case that modifying interpretation of intent might lead to changes in interpretation of facial expression and vice versa, or it could be the case that both of these biases should be trained explicitly simultaneously to maximize the change in (non)hostile

interpretation bias. This is a relevant question because understanding the factors that influence the training effects might provide cues as to how the training might be strengthened.

The experiment reported in **Chapter 4** extends the findings from the study described in **Chapter 2** by examining the effects of modifying interpretation bias of intent using CBM-I paradigms on how participants would interpret ambiguous facial expressions. We expected that the increase in prosocial intent attribution bias in the positive training condition would lead to an increase in prosocial interpretation bias of facial expressions. On the other hand, we expected that the increase in hostile intent attribution bias in the negative training condition would increase hostile interpretation bias of facial expressions.

**Attention Bias in Aggression and Its Modification (CBM-A)**

While it has been suggested that processing social information in a hostile way may be due to deficits in the first step of the SIP model (encoding social cues) (Horsley, de Castro, & van der Schoot, 2010), this step has received only limited attention in experimental studies. Encoding refers to the process of attending (i.e., paying attention) to relevant social cues and placing those cues in the memory for further processing (Brown & Craik, 2000). Interestingly, in the literature, two conflicting hypotheses regarding attentional deployment in relation to aggression can be found (de Castro & van Dijk, 2017). The first hypothesis proposes that aggressive individuals tend to show heightened attention for hostile versus non-hostile social cues (Crick & Dodge, 1994). The emotional Stroop task and the dot-probe are among the most common behavioral paradigms that were used to assess selective attention bias. Typically, in the dot-probe task, participants are presented with either a hostile or a non-hostile word or image, one of which is replaced with a dot. Participants are asked to indicate the location of the dot as quickly as possible by clicking the up and down button. On the other hand, in the emotional Stroop task, participants are presented with words (i.e., aggressive, positive, or negative emotion words) with different font colors. Participants are asked to ignore the emotional content of the word and only report the font color of the word. A number of studies using these assessment tasks found that when participants were presented with both non-hostile and hostile stimuli (e.g., words or images), aggressive participants tended to pay more attention to hostile stimuli and took longer to name the colors of aggressive and negative words (e.g., Smith & Waterman, 2003; Smith & Waterman, 2004; Dodge & Price, 1994).

The second hypothesis proposes that aggressive individuals do not necessarily show heightened attention to hostile versus non-hostile cues; however, they selectively encode (hostile) cues in a way that fits a hostile schema (Horsley et al., 2010; Troop-Gordon, Gordon, Vogel-Ciernia, Lee, & Visconti, 2018). Two recent eye-tracking studies measured participants' selective attention bias toward hostile cues in a sample of aggressive children. Participants' eye movements were recorded in real-time using eye-tracking technology while viewing pictures or video clips of ambiguously hostile situations in which one person is harming another person, but it was unclear whether this harm was intentional.

It was found that aggressive and nonaggressive children did not differ in their attention to hostile and non-hostile cues. However, although aggressive children attended equally to non-hostile cues, they recalled less of those cues, and they were better able to recall hostile cues that were more consistent with their pre-existing hostile schema (Horsley et al., 2010; Troop-Gordon et al., 2018). Also, it was found that aggressive children take longer before fixating on the relevant social cues of the situation (Troop-Gordon et al., 2018). The latter hypothesis specifically could provide important targets to training programs that would not only train aggressive individuals to simply attend to non-hostile rather than hostile cues but also to effectively attend to and encode the most adaptive and relevant social cues that help disambiguate the situation.

The most used CBM-A approach was introduced in anxiety research by MacLeod et al. (2002). It involves using a modified dot-probe task to experimentally induce different attentional responses to a threatening stimulus. In this training, which involves many experimental trials, participants were presented with pairs of words or images that each included one threatening stimulus or one non-threatening stimulus. Participants had to indicate the location of the dot as quickly as possible by clicking the up and down button, which appeared in the locus of either stimuli depending on the training condition. In the training condition that aimed to reduce selective attention to threat, the probe appeared in the opposite locus from the threat stimulus, and in the training condition that aims to increase attention selectivity to threat, the probe appeared in the opposite locus of the neutral stimulus. The study showed that the dot-probe can successfully train attention selectivity to produce an attention bias toward threat cues with an associated increase in stress reactivity and train attention bias toward non-threat cues with an associated decrease in stress

reactivity. This foundational CBM-A study opened the gateway to examine the impact of CBM-A training in a wide variety of other conditions.

Studies have shown that manipulation of attention bias was successful not only in improving symptoms of anxiety and stress reactivity (see Bar-Haim, 2010, for review) but also social phobias (e.g., Amir et al., 2009), chronic pain syndrome (e.g., McGowan, Sharpe, Refshauge, & Nicholas, 2009), depression (see Hallion & Ruscio, 2011, for meta-analytic review), body dysmorphia (e.g., Smeets, Jansen, & Roefs, 2011), and alcohol dependency (e.g., Schoenmakers et al., 2010). However, an important challenge of applying this training methodology in the context of aggression is that task features related to the dot-probe (i.e., inferred focus) would not be able to properly target the nature of attention bias in aggression. As mentioned earlier, aggression is associated with the necessity of a longer time to attend to relevant social cues (Troop-Gordon et al., 2018) and with selectively encoding cues (i.e., hostile cues) that fit a hostile interpretation (Horsley et al., 2010; Troop-Gordon et al., 2018). In this case, probe-based CBM-A training programs might not be the most optimal procedure for modifying gaze patterns associated with aggression. Thus, there is a need for training programs to train precise attention components implicated in aggression to meet the unique needs of aggressive individuals.

A number of eye-tracking studies provided encouraging results for a novel training methodology implementing gaze-contingency. It shows potential for not only modifying attentional selectivity but also for its potential clinical utility. Gaze-contingency is an online interactive technique that allows the computer screen display to change based on where the individual is looking in real-time via eye-tracking technology (Wang et al., 2015). The major advantage of this procedure is that the setup enables direct assessment and training of gaze direction, unlike indirect probe-based CBM-A training paradigms that target only the end of an attentional process (Lazarov, Pine, & Bar-haim, 2017; Price, Greven, Siegle, & Koster, 2016). Recent studies in the context of depression and anxiety show that attention can indeed be trained successfully using gaze-contingency techniques (Ferrari, Mobius, van Opdorp, Becker, & Rinck, 2016; Lazarov et al., 2017; Price et al., 2016). For instance, Lazarov et al. (2017) trained anxious participants using gaze-contingent music reward therapy in order to reduce attention-dwelling on threat stimuli associated with social anxiety disorder. Participants had to fix their attention on the neutral stimuli (i.e., neutral facial expressions) when presented with other facial expressions in

order for the music of their choice to play; if the participant attended to threat stimuli (i.e., disgusted facial expressions), then the music stopped. The training resulted in reduction in self-reported, clinical-rated anxiety and in dwell-time on threat stimuli.

Interestingly, however, we are not aware of any study in aggression that has sought to train attention bias using gaze-contingencies. Furthermore, it would be of great interest to examine the effects of training paradigms on adaptive social stimuli, attention bias, and aggression. Paradigms that train individuals to maintain gaze could, in principle, provide the greatest benefits in reducing aggressive behavior and provide an interesting avenue for future intervention research in the context of aggression.

The experiment presented in **Chapter 3** provided a first step in aggression studies toward the development of attention bias training using a novel gaze-contingent CBM-A procedure. Male and female university students were trained to attend to either adaptive or maladaptive cues. Effects on attention bias, aggression (self-reported and behavioral measure), anger, and mood were assessed. We predicted that training individuals to attend to adaptive cues would increase adaptive attention and might reduce subsequent aggressive behavior. On the other hand, training them to attend to maladaptive cues would increase maladaptive attention and increase subsequent aggressive behavior.

**Combining CBM-I and CBM-A Approaches**

There is an emerging experimental interest in the potential intervention value of delivering both CBM-A and CBM-I in combination. The first to advance this notion were Hirsch, Clark, and Mathews (2006), who formulated the combined cognitive bias hypothesis. This hypothesis states that: “Cognitive biases do not operate in isolation, but rather can influence each another and/or can interact so that the impact of each on another variable is influenced by the other. Via both these mechanisms we argue that combinations of biases have a greater impact on disorders than if individual cognitive processes acted in isolation” (p. 224). Experimental studies that examined the combined cognitive bias hypothesis tested it by delivering both CBM-A and CBM-I training in combination. Suggesting that training procedures that target a combination of biases have a greater impact on disorders than targeting a cognitive process in isolation. For example, Brosan et al. (2011) confirmed the effectiveness of combined attention and interpretation bias training in reducing attention bias to threat and increasing positive interpretation bias. Additionally, the

combined training led to a reduction in state and trait anxiety in a sample of anxious outpatients. Moreover, Beard, Weisberg, and Amir (2011) provided evidence that a combined CBM-I and CBM-A can significantly reduce anxiety symptoms in patients with social anxiety disorder compared to a control group, and the reported intervention effect of the combined CBM was moderate to large.

Although the combined cognitive hypothesis focused on cognitive processes in social anxiety, the hypothesis might also be applicable to other clinically relevant conditions. This is especially true when we consider that the SIP model postulates that cognitive biases such as attention and interpretation in aggression are associated rather than independent (Crick & Dodge, 1994). However, before examining the effect of CBM-I and CBM-A in combination, an important starting point is to better understand the interactive effects between attention and interpretation bias in the context of aggression. This knowledge is relevant; if cognitive biases of attention and interpretation influence one another and interact in maintaining aggression, then targeting both biases in combination may potentially maximize aggression reduction. We are not aware of any aggression studies that have examined the interrelation between these biases using CBM paradigms. Additionally, the current research that has been done has mostly studied cognitive biases in isolation, since this single approach enhances our understanding of how a specific cognitive bias affects aggression. However, it is limited as it does not provide insight as to how cognitive biases are associated and how various biases may influence the etiology and maintenance of aggression. Especially since previous anxiety research has indicated that modifying one bias may have an indirect effect on other biases (Amir et al., 2010; Hirsch et al., 2006; White, Suway, Pine, Bar-Haim, & Fox, 2011). For example, White et al. (2011) designed a CBM-A training procedure to induce attention bias to threat and examine its effect on interpretation to an anxious sample. The results indicated that individuals who participated in the attention to threat manipulation training showed an increase in anxiety-related negative interpretations of ambiguous situations compared to the placebo training group. Also, Amir et al. (2010) found that CBM-I was successful not only in modifying interpretations in a socially anxious sample but also in influencing attention biases to threat stimuli.

Additionally, examining the effects of attention training on interpretation in aggression (and vice versa) would provide a better understanding of how attention and interpretation biases interact

and contribute to aggressive behavior. For example, in an anxious sample, Bowler et al. (2017) trained one group using the CBM-I paradigm and the other group using the CBM-A paradigm. The results showed that while CBM-A was successful in transferring the effect of the modified attentional bias to subsequent changes in interpretation bias, CBM-I failed in modifying subsequent attentional bias. These findings suggest that compared to CBM-I, CBM-A may have more of a generalizable cognitive effect. In the context of aggression, it is possible that focusing on one cognitive bias may be insufficient to cause change in another bias and impact aggression, especially compared to a combined training that includes a combination of biases that might have a greater impact on reducing these biases and aggression. Regardless of whether future studies support isolated or combined cognitive bias training, the results will undoubtedly provide new directions for further development in CBM techniques in reducing aggressive behavior.

As a first step, the study described in **Chapter 3** investigated the interrelation between attention and interpretation bias, by examining the effects of CBM-A on how subsequent ambiguous social information was interpreted. We expected that participants who were trained to attend to adaptive cues would make less hostile interpretations than participants who were trained to attend to maladaptive cues. Next, the study described in **Chapter 4** extended the findings of the possible interrelation between attention and interpretation bias, by examining the effects of the modified interpretation bias of intent on attention bias. We expected that the increase of prosocial interpretation bias of intent would lead to heightened attention to adaptive cues, and that the increase in hostile interpretation bias of intent would lead to heightened attention to maladaptive cues. Finally, the experiment presented in **Chapter 5**, which was built on experiments from **Chapter 3** and **Chapter 4**, investigated the effect of a combined CBM-A and CBM-I training paradigm on modifying both interpretation and attention bias and explored the effects of this manipulation on aggression. We expected that a combined training program would have stronger effects on reduction of aggression than training attention and interpretation biases in isolation.

**Focus and Research Questions of This Dissertation**

Aggression studies are limited in examining the effects of CBM-I, and there have been no studies on the effectiveness of CBM-A on both attention bias and aggression reduction. In addition, training paradigms in this area typically assess and train using written vignettes (e.g., Hawkins & Cougle, 2013; Vassilopoulos et al., 2015). However, in real-life situations, visual nonverbal cues

such as facial and physical expressions carry important signs regarding the intentions of others (Cadesky et al., 2000). Therefore, the general aim of the present dissertation is to examine whether novel CBM-A and CBM-I procedures using pictorial stimuli can be used to change maladaptive information processing in the context of aggression. Most importantly, we want to examine the effects of the altered aggression-related cognitive biases on concrete aggressive behavior using self-report and behavioral measures. Additionally, the previous literature suggests that cognitive biases, such as attention and interpretation in aggression, are associated rather than independent (Crick & Dodge, 1994; Hirsch et al., 2006). Therefore, we aim to explore how attention and interpretation biases interact in maintaining aggression. Further, in line with studies that suggest that training procedures that target a combination of biases have a greater impact on symptom reduction than targeting cognitive processes in isolation (Hirsch et al., 2006), we aim to establish whether this novel CBM paradigm should target attention, interpretation, or both for the best results. Given the relative scarcity of CBM studies in the context of aggression to date and the novelty of our training procedure, we aim to examine our novel training procedure on both biases and aggression in an unselected sample of students. This would make it possible to first draw conclusions regarding the possible effects of such a training procedure on both biases and aggression before applying our training procedure to a clinical sample.

The current dissertation focused on four questions:

1- Can a novel CBM training procedure using pictorial stimuli be used to change interpretation and attention biases in the context of aggression?

2- Do changes in attention or interpretation biases lead to changes in aggression?

3- How do attention and interpretation biases interact in maintaining aggression?

4- Is a combined bias CBM training procedure more effective than a single bias CBM training procedure on both bias and aggression reduction?

Given the fact that we used pictorial stimuli to train participants based on the idea that facial expressions contain information regarding intentions of others, an additional question of concern was whether changes in attribution bias of intent affects interpretation bias of facial expressions.

To answer the research questions, four studies are included in this dissertation and are described in more detail in the upcoming chapters (**Chapter 2** to **6**). Below we provide a short outline of the dissertation.

As a first step, **Chapter 2** describes a pilot study which examines the effects of a novel CBM-I training procedure using pictorial stimuli on modifying interpretation bias and aggression. Next, **Chapter 3** examines the efficacy of a novel gaze-contingent CBM-A procedure on modifying attention bias and aggression. Additionally, the chapter addresses how attention and interpretation bias interact by examining the effect of modifying attention on interpretation bias. **Chapter 4** extends the findings of **Chapter 1** by addressing how attention and interpretation bias interact by examining the effect of modifying interpretations on attention bias. Additionally, the study in this chapter further examines the effects of modifying interpretations of intent on interpreting ambiguous facial expressions. **Chapter 5** takes the next step and explores the effect of a combined CBM-A and CBM-I training paradigm on modifying interpretation and attention biases and examines the effects of this manipulation on aggression. Finally, **Chapter 6** provides a summary and general discussion of the main findings of this dissertation.

# CHAPTER 2

## The Effects of a Novel Hostile Interpretation Bias Modification Paradigm on Hostile Interpretations, Mood, and Aggressive Behavior

**This chapter has been published as:**
AlMoghrabi, N., Huijding, J., & Franken, I. H. (2018). The effects of a novel hostile interpretation bias modification paradigm on hostile interpretations, mood, and aggressive behavior. *Journal of Behavior Therapy and Experimental Psychiatry, 58*, 36-42.

**Abstract**

*Background and objectives*: Cognitive theories of aggression propose that biased information processing is causally related to aggression. To test these ideas, the current study investigated the effects of a novel cognitive bias modification paradigm (CBM-I) designed to target interpretations associated with aggressive behavior.

*Methods*: Participants aged 18–33 years old were randomly assigned to either a single session of positive training ($n = 40$) aimed at increasing prosocial interpretations or negative training ($n = 40$) aimed at increasing hostile interpretations.

*Results*: The results revealed that the positive training resulted in an increase in prosocial interpretations while the negative training seemed to have no effect on interpretations. Importantly, in the positive condition, a positive change in interpretations was related to lower anger and verbal aggression scores after the training. In this condition, participants also reported an increase in happiness. In the negative training no such effects were found. However, the better participants performed on the negative training, the more their interpretations were changed in a negative direction and the more aggression they showed on the behavioral aggression task.

*Limitations*: Participants were healthy university students. Therefore, results should be confirmed within a clinical population.

*Conclusions*: These findings provide support for the idea that this novel CBM-I paradigm can be used to modify interpretations, and suggests that these interpretations are related to mood and aggressive behavior.

Research into the social cognitive aspects of aggressive behavior has shown that aggressive individuals frequently display cognitive biases in the processing of environmental stimuli (Quiggle, Garber, Panak, & Dodge, 1992). According to the social information processing (SIP) model (Crick & Dodge, 1994), an individual's social behavior is a function of six steps: (1) encoding of social cues; (2) interpretation of those cues; (3) setting goals; (4) formulating responses; (5) evaluating different responses until an acceptable response is generated; and (6) response enactment. Adequate processing of social information during these steps will lead to adaptive behaviors, while biased processing may result in maladaptive behaviors, including aggression.

In line with this model, reactive aggression has been found to be associated with biases in encoding and interpreting social cues (e.g., Dodge, 2006). With respect to the interpretation of social cues, a meta-analytic review found that more hostile attributions are strongly related to more aggressive behavior (Orobio de Castro, Veerman, Koops, Joop, & Monshouwer, 2002). For example, Crick and Dodge (1996) showed in a sample of aggressive and non-aggressive children aged nine to 12 that reactive aggressive children more often attributed hostile intent to peers than non-aggressive children and that these hostile attributions motivated aggressive behavior. Such findings inspired the development of a number of interventions aimed at preventing or reducing aggressive behavior by manipulating social information processing.

One way to manipulate social information processing is by employing cognitive bias modification (CBM). This paper focuses on the effects of manipulating interpretation bias (CBM-I) on aggression. Such CBM-I procedures are designed to modify interpretations of the intentions of others, by exposing participants multiple times to ambiguous social situations and training them to interpret these situations either in a negative (i.e., hostile) or positive (i.e., prosocial) way using feedback. For example, Vassilopoulos, Brouzos, and Andreou (2014) trained a sample of 10–12-year-old children using a three-session attribution training program, and found that hostile attributions regarding ambiguous social situations decreased while positive attributions increased.

Studies in adult samples have also suggested that hostile interpretations can be modified using CBM procedures (Hawkins & Cougle, 2013; Penton-Voak et al., 2013). For example, Hawkins and Cougle (2013) randomly assigned a number of undergraduate students to a positive training, a negative training, or a control condition. The positive training led to an increase in positive interpretation bias whereas the negative training led to an increase in negative

interpretation bias. Importantly, participants in the positive training also reported less angry responses in reaction to an insult than participants from the other conditions.

Although the results of these first studies on the effects of CBM-I on aggression are promising, there is a dire need for studies replicating and extending these initial promising results.

The current study aimed to replicate the finding that interpretational styles can be altered and that this impacts aggression, using a new CBM-I paradigm that includes visually rather than verbally presented ambiguous social situations. In real-life situations, visual nonverbal behaviors (e.g., facial and physical expressions) hold important social information about the internal state (including intentions) of the other person (Cadesky, Mota, & Schachar, 2000). Indeed, research has shown that aggressive children inaccurately interpret cues of benign and prosocial intention as hostile (Dodge, Murphy, & Buchsbaum, 1984). This suggests that including visual ambiguous social scenes, rather than written stories (i.e., vignettes), might boost the effects of the training procedure. Based on previous studies (e.g., Hawkins & Cougle, 2013; Penton-Voak et al., 2013), we expected that training individuals to interpret ambiguous situations as non-hostile would lead to a reduction in aggressive behavior whereas training them to interpret such situations as hostile would increase aggressive behavior. Given that previous findings show that manipulating interpretation bias can also impact mood (e.g., Lothmann, Holmes, Chan, & Lau, 2011), we also included measures of mood before and after the training.

## Method

### *Participants*

Forty male and forty female students from Erasmus University Rotterdam (42 Caucasians, 12 Asian, 6 Middle Eastern, 4 Hispanic, 1 African, and 15 others), aged between 18 and 33 ($M$ = 21.67, $SD$ = 3.17) participated in exchange for course credits.

### *CBM-I Training*

The training task consisted of 52 trials that were presented using E-prime software. For each trial, participants viewed a different image of a hypothetical social situation in which one person harmed another. These images were used to assess and manipulate interpretation bias. The training task was completed within a single session and consisted of three phases: baseline, training, and test. The baseline and test phases consisted of six trials during which interpretation bias was assessed. The training phase consisted of forty trials during which interpretations were

manipulated. Participants were randomly assigned to the positive or the negative training condition.

Phase 1 (baseline) and 3 (test): On each trial participants were presented on the computer screen with a single sentence scenario that described a negative situation. For example, “His arm bumped hard into him!” Participants were then presented with an image of a social situation in which a mishap occurred which was ambiguous with respect to the intent of the harm-doer (see Figure 1). After 200 ms, two rectangles appeared on the image, one around the face of the harm-doer and the other around the focus of the incident (e.g., the place where the “victim” is hit by the arm). Participants were first asked to click on the rectangle surrounding the place in the picture that best indicated whether or not the mishap occurred on purpose. We included this assessment to get an idea of what kind of information in the scene would be deemed most important by participants for disambiguating the situation. A discussion of these exploratory data are beyond the scope of the current manuscript. Thereafter, the question “Why did this happen?” along with two possible interpretations, one hostile and one benign, appeared on the screen. For example, the picture presented in Figure 1 was accompanied by the following two interpretations: (a) This happened on purpose because he doesn’t want him to pass (hostile interpretation); (b) This happened by accident because he didn’t see him (non-hostile interpretation). Participants were asked to rate for each interpretation how likely they considered it to be true, by marking a 100 point visual analogue scale that was anchored with the labels “No, definitely not” on the left and “Yes, definitely” on the right ends.

Phase 2 (training): On each trial participants were presented with an image of a social situation in which a mishap occurred, which was ambiguous with respect to the intent of the harm-doer. The images were always preceded by a short description of the situation. All scenarios were one sentence long, and described the negative outcome. For example, the image presented in Figure 2 was preceded by the description: “His drawing is all ruined!” The image was presented on the screen until the spacebar was pressed, after which the question “Why did this happen?” appeared on the screen. After clicking the mouse to continue, a hostile and one non-hostile interpretation appeared simultaneously on the screen, randomly positioned one above the other. Participants were asked to click on the interpretation they considered to be most likely. In the positive training condition, the non-hostile interpretations were reinforced as “correct” while, in the negative training, the hostile interpretations were “correct”. For example, the situation depicted

in Figure 2 was accompanied by the following two interpretations: (a) “This happened on purpose because he dislikes him”; (b) “This happened by accident because he bumped against him” Following a “correct” response, the word “CORRECT” was presented at the top of the screen in green font, the color of the font of the selected interpretation and the line around it changed from navy blue to green, and the other interpretation disappeared to avoid confusion regarding the feedback. Following an “incorrect” response, the word “INCORRECT” was presented at the top of the screen in red font, the color of the font of the selected interpretation and the line around it changed from navy blue to red, and the other interpretation then disappeared from the screen. Feedback remained on the screen for 2,000 ms, after which the next trial began.

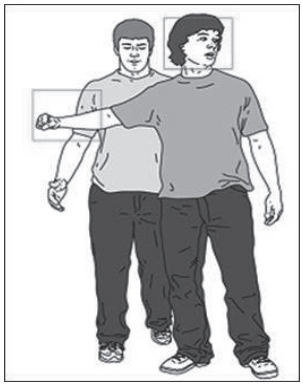


*Figure 1.* Example from the baseline phase.

*Figure 2.* Example from the training phase.

*Stimulus materials*

A set of 52 pictures were used to assess and train interpretation bias. Each image depicted a situation in which one person harmed another. For the baseline and test phases we used images from the study of Wilkowski, Robinson, Gordon, and Troop-Gordon, (2007; see Figure 1). For the training phase, we used images from the study of Horsley, de Castro, and Van der Schoot (2010; see Figure 2), supplemented by thirty images from stock image websites. The pictures were selected to vary in their level of ambiguity regarding the intent of the harm-doer, just like the types of situations we encounter in day-to-day life, but should not provide clear cut cues on intentionality. Thus for each picture it should be the case that the harm could in principle be either intentional or unintentional.

To evaluate the adequacy of the stimulus materials a pilot test was carried out. Forty university students were asked to rate the pictures on a number of characteristics, including the extent to which the depicted harm was intentional. Intentionality was rated on a 100 point VAS scale that was anchored with the labels "Accidental" on the left and "Intentional" on the right ends. The results show that the pictures were rated on average as very ambiguous for the baseline and test phase $M = 51.3$, $SD = 14.1$, range = 20.8 – 81.7, as well as the training phase $M = 47.0$, $SD = 11.6$, range = 16.2 – 69.2. Thus, the intentionality ratings of the pictures varied within and between pictures, indicating that they were indeed ambiguous with respect to the intent of the harm-doer.

*Measures*

*Aggression Task*

Aggression was measured post-training using the Taylor Aggression Paradigm (TAP; Taylor, 1967). Participants were told that they would be competing against an opponent on a competitive reaction time game consisting of 25 trails. Depending on whether they won or lost a trial, they would either receive a noise blast from the opponent or be allowed to administer a noise blast to the opponent. The experiment was presented as a collaboration between Erasmus University and Utrecht University for which the opponent was currently present at a lab in Utrecht receiving the same instructions. In reality no experimental collaboration or opponent existed, and the arrangement of winning and losing on each trial as well as the level of noise administered by the opponent was pre-programmed (see Appendix; cf. Brugman et al., 2014). Each participant was

seated at a table facing a computer screen and a mouse. A message on the screen "Connecting" appeared to have the participant believe that his/her computer was connecting with that of the opponent. Participants were instructed that the aim of the task was to click faster than their opponent on a designated rectangle when it turned from yellow to red. Depending on whether the trial was won or lost the message "You Won" or "You Lost" appeared on the screen, and the winner was supposedly allowed to administer a noise blast to the opponent. Before administering a blast, the participant had to select the duration (between 0 and 10 seconds) and the volume of the noise (between 0 and 100 dB). After losing a trial, the participant received a noise blast through the headphones and were given feedback regarding the level and duration of that noise.

*Questionnaires*

In order to assess state aggression prior to the training, we reworded Buss and Perry's (1992) trait Aggression Questionnaire (AQ) following the same method used by Farrar and Krcmar (2006). The adapted questionnaire started with the following instruction: "Imagine that you just bought something to drink. When you walk outside, somebody bumps into you, spilling your drink over your favorite clothes. As you look at the mess, you hear this person swearing." Then followed 20 items from the AQ that were reworded to describe possible reactions to the abovementioned situation. For example, the original AQ item "Sometimes I fly off the handle for no reason" was reworded to "I might fly off the handle for no reason with this person" to reflect state aggression. Participants rated how characteristic each response would of them on a 7-point scale (1 = extremely uncharacteristic; 7 = extremely characteristic). The questionnaire consisted of three subscales: physical aggression, verbal aggression, and anger. After the training, participants completed the same items but with a different story: "Imagine that you are at the Starbucks working on an assignment. Suddenly, someone bumps into your table, spilling coffee all over your notes. You see that the other person looks really annoyed." In our sample, Cronbach's alphas were .93 and .92 for the pre- and post-assessments, respectively.

The Reactive-Proactive Aggression Questionnaire (RPQ; Raine et al., 2006) was administered to assess reactive (11-items) and proactive (12-items) aggression on a 3-point scale (0 = Never, 1 = Sometimes, and 2 = Often). In our sample, Cronbach's alpha was .77.

Part B of the Novaco Anger Scale (NAS; Novaco, 1994) was administered to measure anger intensity across 25 potentially provoking situations, on a 5-point scale from 0 (no annoyance) to 4 (very angry). In our sample, Cronbach's alpha was .90.

To assess mood, participants indicated how happy, angry, sad, and afraid they felt at that moment by marking visual analogue scales that were anchored with the labels "not at all" on the left and "very much so" on the right ends. In addition, participants completed the 20-item Positive Affect and Negative Affect Schedule (PANAS; Watson et al., 1988), consisting of 10-negative and 10-positive affective states which are rated on the extent to which they apply to the participant "right now", on a five point scale (1 = Slightly; 5 = Extremely). In our sample, Cronbach's alpha was .79.

For exploratory purposes beyond the scope of this manuscript the State-Trait Anxiety Inventory (STAI) was also included (Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983).

*Procedure*

After receiving instructions and completing an informed consent, participants completed the AQ, STAI, RPQ, NAS questionnaires, and the mood VASs. They then began the CBM-I training, followed by the mood VASs, the TAP, the AQ and the PANAS.

## Results

*Data reduction and preliminary analyses*

First we calculated interpretation bias (IB) scores for the pre- and post-treatment assessments by subtracting the mean VAS truth rating for the negative interpretations from the mean VAS truth rating for the positive interpretations. Thus, positive IB scores indicate that positive interpretations were rated as more likely to be true than the negative interpretations.

Next, in order to ascertain the appropriateness of our IB measure, we correlated the interpretation bias scores (IB-pre and IB post) and the concurrently assessed aggression outcome measures. IB scores correlated significantly with concurrent AQ scores before ($r$ = -.28, $p$ = .011) and after the training ($r$ = -.27, $p$ = .016), specifically with the verbal (pre: $r$ = -.34, $p$ = .002, post: $r$ = -.25, $p$ = .024) and the anger (pre: $r$ = -.35, $p$ = .002, post: $r$ = -.29, $p$ = .010) subscales. In addition, IB scores after the training correlated significantly with the TAP scores (total: $r$ = -.30, $p$ = .008; intensity: $r$ = -.32, $p$ = .004; duration $r$ = -.32, $p$ = .004). This provides some support for

the validity of our approach as it shows that we assessed and trained interpretations that are meaningfully related to aggression.

Finally, to get an idea of whether the novel training approach was clear and doable for participants, we explored participants' accuracy during training. While participants in the positive training made few errors ($M$ = 17.6%, $SD$ = 9.86), this was not the case in the negative condition, in which significantly more errors were made ($M$ = 51.6%, $SD$ = 20.42, $t(78) = -9.71$, $p < .01$).

*Baseline measures*

Independent-samples t-tests confirmed that the positive and negative training groups did not differ significantly in the baseline levels of self-reported aggressive behavior (AQ and RPQ), anger (NOVACO), anxiety (STAI-ST), and mood ratings (happy, angry, sad, and afraid). Descriptive statistics for the pre-training measures are presented in Table 1. In addition, the groups did not differ significantly in their interpretation bias prior to the training: all $t$ values < 1.21; all $p$-values > .227.

Table 1

*Descriptive Statistics for Pre- and Post-Training Measures*

| Measures | Positive training | | Negative training | |
|---|---|---|---|---|
| | *M* | *SD* | *M* | *SD* |
| Pre-training | | | | |
| Aggression Questionnaire | 64.15 | 19.70 | 64.50 | 20.13 |
| Physical Aggression | 25.03 | 9.64 | 26.45 | 9.46 |
| Verbal Aggression | 16.77 | 4.99 | 16.40 | 5.94 |
| Anger | 22.35 | 7.90 | 21.65 | 7.07 |
| Reactive-Proactive | 31.15 | 4.56 | 31.70 | 4.10 |
| NOVACO Anger Scale | 67.23 | 13.89 | 66.22 | 14.21 |
| Anxiety Inventory-State | 35.68 | 8.91 | 34.13 | 8.92 |
| Anxiety Inventory-Trait | 44.52 | 11.94 | 40.03 | 8.82 |
| Angry mood | -39.23 | 17.63 | -39.70 | 16.14 |
| Afraid mood | -42.32 | 14.42 | -42.50 | 15.34 |
| Sad mood | -27.87 | 25.23 | -27.85 | 25.31 |
| Happy mood | 15.67 | 22.74 | 19.60 | 16.57 |
| Post-training | | | | |
| Aggression Questionnaire | 62.48 | 20.70 | 65.00 | 21.43 |
| Physical Aggression | 24.83 | 9.18 | 27.52 | 10.69 |
| Verbal Aggression | 16.08 | 5.49 | 15.78 | 6.62 |
| Anger | 21.58 | 8.13 | 21.70 | 6.68 |
| PANAS-positive | 29.55 | 6.66 | 30.18 | 7.21 |
| PANAS-negative | 21.87 | 5.34 | 22.85 | 5.93 |
| Angry mood | -39.03 | 15.29 | -34.33 | 19.69 |
| Afraid mood | -41.20 | 13.90 | -42.08 | 12.21 |
| Sad mood | -32.93 | 22.50 | -30.43 | 23.20 |
| Happy mood | 20.10 | 21.02 | 18.37 | 17.89 |
| Taylor Aggression Paradigm | 19.12 | 15.04 | 21.50 | 19.02 |

*Effects of training on interpretation bias*

To examine the effects of training on interpretation bias, the IB scores were subjected to a 2 Assessment (pre, post-treatment) x 2 Group (negative versus positive training) ANOVA with repeated measures. The analysis revealed significant main effects for the group, $F(1, 78) = 4.68$, $p = .033$, $\eta_p^2 = .06$, and the assessment, $F(1,78) = 18.35$, $p < .001$, $\eta_p^2 = .19$. More importantly, the crucial interaction between the group and the assessment was significant, $F(1, 78) = 11.52$, $p = .001$, $\eta_p^2 = .13$ (see Figure 3). This interaction was decomposed using paired-samples t-tests. This showed that in the positive condition, interpretation bias became significantly more positive: $t(39) = -7.01$ $p < .001$. In the negative condition, interpretation bias scores did not change significantly over time: $t(39) = -.53$, $p = .598$.

To explore whether the accuracy during training could have influenced the effects of the training on changing interpretations, we calculated interpretation bias change scores by subtracting the IB score before the training from the IB score after the training. Thus, more positive IB change scores indicate that participants interpretations of the situations became more positive (i.e., prosocial). In the negative condition the change in interpretation bias was significantly correlated with participant's accuracy scores ($r = -.52$, $p < .001$). Perhaps not surprisingly given the lower variability in accuracy rates, this effect was less strong in the positive condition ($r = .27$, $p = .098$).

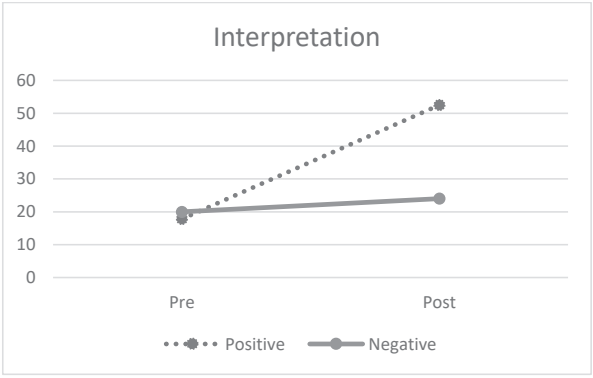



*Figure 3.* Average interpretation ratings at pre- and post-training for each training condition.

*Effects of interpretation training on aggression*

Aggression scores from the AQ were subjected to a 2 Assessment (pre, post-treatment) x 2 Group (negative versus positive training) ANOVA with repeated measures. The analysis revealed no main effects of the group or the assessment and no significant interaction between the group and assessment, $F(1, 78) = 1$, $p > .321$ (see Figure 4).

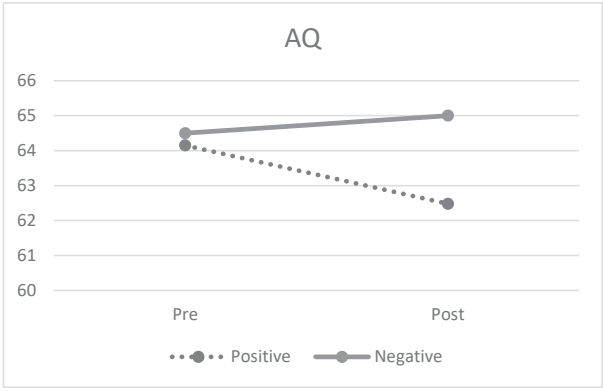



*Figure 4.* Average Aggression Questionnaire (AQ) ratings at pre- and post-training for each training condition.

Additionally, an independent-samples t-test showed no group differences in TAP performance ($t(78) = 0.62$, $p = .537$), intensity ($t(78) = 0.80$, $p = .429$), and duration ($t(78) = -0.28$, $p = .781$).

Given the novelty of the training task, we additionally performed a number of exploratory analyses. First, while the training did not result in changes in our primary outcome measures at the group level, it is possible that the impact of the training varied between individuals and that the extent to which the training successfully changed interpretations. To explore this possibility, we correlated the IB change score with various outcome measures. The change in interpretation bias within the positive condition showed a significant negative correlation with the post-training AQ total score ($r = -.34$, $p = .032$) and with the anger ($r = -.33$, $p = .037$) and verbal ($r = .34$, $p = .005$) subscales. This suggests that the more the interpretation bias changed in a pro-social direction, the less anger and verbal aggression participants reported after the training. A significant negative correlation between the interpretation bias change score and the AQ verbal subscale before the

training ($r$ = -.36, $p$ = .022) suggests that it is also possible that those participants who reported being less verbally aggressive were more likely to benefit from positive interpretation bias training. However, the change in interpretations was not significantly related to the (pre-training) RPQ-proactive ($r$ = .05, $p$ = .77) and RPQ-reactive ($r$ = -.17, $p$ = .298) scores, indicating that changes in interpretations during the positive training were independent of prior levels of reactive and proactive aggression. Unsurprisingly, given the overall lack of change in the interpretation bias scores in the negative condition, the change in interpretations within the negative condition did not correlate significantly with the post-training AQ scores ($r$ = .07, $p$ = .654) or its subscales. In addition, the change in interpretations in the negative condition was not significantly related to the RPQ-proactive ($r$ = -.17, $p$ = .289) and RPQ-reactive ($r$ = -.01, $p$ = .949) scores. Furthermore, the IB change score was not significantly related to the TAP scores in either the positive condition in general ($r$ = -.15, $p$ = .346), in terms of intensity ($r$ = -.11, $p$ = .499), or duration ($r$ = -.20, $p$ = .220), or the negative condition in general ($r$ = -.02, $p$ = .886), in terms of intensity ($r$ = -.08, $p$ = .624), and duration ($r$ = -.05, $p$ = .773).

Secondly, we explored the influence that training accuracy may have had on the effects of training on the outcome measures. Therefore we correlated participants' accuracy during the training with various outcome measures. Accuracy did not correlate significantly with the post-training AQ scores either in the positive ($r$ = .15, $p$ = .368) or the negative condition ($r$ = .15, $p$ = .360, respectively). The same was true for the correlations with the AQ subscales.

However, accuracy was significantly related to aggressive responding on the TAP. That is, the better the participants performed during the negative training the more aggressive their responses on the TAP in general ($r$ = .32, $p$ = .044), intensity ($r$ = .42, $p$ = .007) and duration ($r$ = .39, $p$ = .014). This suggests that the negative training did have an effect on those participants who performed well. In the positive group, the accuracy during training was not significantly related to the TAP scores. This latter finding was not very surprising since the participants in the positive condition uniformly made very few errors.

*Effects of interpretation training on mood*

VAS mood ratings (happy, angry, sad, and afraid) were subjected to separate 2 Assessment (pre, post-treatment) x 2 Group (negative versus positive training) ANOVAs with repeated measures. The analyses revealed that only for self-reported happiness the crucial Assessment x

Group interaction was significant, $F(1, 78) = 4.45$ $p = .038$, $\eta_p^2 = .05$. This interaction was decomposed using a paired-samples t-tests. This showed that in the positive condition, there was a significant increase in self-reported happiness from pre- to post-training, $t(39) = -2.50$, $p = .018$, while in the negative condition, there were no significant changes in happiness from pre- to post-training, $t(39) = .62$, $p = .542$. For self-reported anger the crucial Assessment x Group interaction showed a trend towards significance, $F(1, 78) = 3.01$ $p = .086$, $\eta_p^2 = .04$. Explorative paired-samples t-tests showed that in the positive condition, there were no significant changes in self-reported anger from pre- to post-training, $t(39) = -.10$, $p = .924$, while in the negative condition, there was a significant increase in self-reported anger from pre- to post-training, $t(39) = -2.51$, $p = .016$.

In addition, the post-training PANAS scores were compared between the two conditions. Independent-samples *t*-tests showed that neither the positive nor the negative affect scores differed significantly between the two conditions.

## Discussion

The current study explored whether a novel cognitive bias modification of interpretation (CBM-I) procedure, designed to modify interpretation bias using pictorial stimuli, influences interpretations and aggressive behavior. The results can be summarized as follows: First, a single session of positive interpretation training using pictorial stimuli resulted in an increase in prosocial interpretation bias. Second, the more the positive training succeeded in changing interpretations in a pro-social direction, the less anger and aggression and more happiness was reported. Third, while a single session of negative interpretation training had no general effect on interpretation, the better participants performed on the negative training, the more their interpretation bias changed. Fourth, the better participants performed on the negative training the more aggressive their responses on a behavioral aggression task.

The current finding that the positive training condition increased prosocial interpretation bias is well in line with previous findings demonstrating that interpretation bias can be trained (Hawkins & Cougle, 2013; Penton-Voak et al., 2013; Vassilopoulos et al., 2014). The finding that participants in the positive training condition also reported a reduction in verbal aggression is interesting since few studies have reported verbal aggression change based on an interpretation intervention. The positive training additionally increased happy mood, which is consistent with

past studies demonstrating that modifying interpretation bias improves mood (e.g., Holmes, Lang, & Shah, 2009; Holmes, Mathews, Dalgleish, & Mackintosh, 2006; Lothmann et al., 2011). It should be noted that, since the negative group did not show a significant decrease in happy mood we cannot rule out the possibility that the significant increase on happy mood in the positive group may be attributed to some other influences. For instance, participants in the positive training were responding more correctly throughout the training compared to participants in the negative training and therefore received more positive feedback which may have influenced mood. However, if the effect of mood was simply due to receiving positive feedback rather than giving a specific response (i.e., selecting a positive) one would expect the accuracy rate to be correlated with positive mood regardless of the experimental condition. This was not the case: the change in happy mood was only related to the accuracy in the positive and not in the negative condition.

The results of the negative training condition on average did not show any change in the participants' interpretation bias. These findings contrast with those of Hawkins and Cougle (2013), which showed that negative training was successful in increasing hostile interpretation bias. A possible explanation is that the current study sample included healthy students compared to the study of Hawkins and Cougle (2013), in which only participants scoring high on trait anger were recruited who may be more susceptible to the effects of a negative training. Interestingly, participants who performed well during the current negative training also showed more change on their interpretation bias. The high number of errors in the negative training seems to suggest that at least part of the participants in the current study actively resisted the negative training by insisting on choosing the benign interpretation despite negative feedback. This may also explain why the negative training did lead to a general increase in the self-reported angry mood from pre- to post-training. It is possible that participants in the negative training were inclined to make prosocial interpretations, and became angry by repeatedly receiving negative feedback. However, the study of Lothmann et al. (2011) have shown that despite that participants in the negative condition made more errors when completing a CBM-I training, the training led to a significant increase in negative interpretation and decrease in positive affect.

Alternatively, and in line with prior studies, the increase in angry mood in the negative condition can be taken as support for the association between hostile interpretations and anger (Wilkowski et al., 2007). However, since the negative training showed no overall significant effect

on participants' interpretation bias, and the change score for the interpretation bias did not correlate with those on the anger mood, it remains unclear whether the interpretation training led to the observed increases in anger levels due to its effects on interpretations or whether the nature of the negative training elicited anger.

While the negative training in general also did not appear to have an effect on the TAP, those participants who performed well on the negative training also showed more reactive aggression on this task. This indicates that training hostile interpretations might have had an effect on aggressive responses, but only to the extent that participants allow themselves to be trained. To our knowledge, few studies have explored the effect of training interpretation bias change on a behavioral aggression task rather than through self-reported measures (e.g., Hawkins & Cougle, 2013). This initial study allowed us to test how the modification of interpretation bias can influence aggressive responses in the context of a competitive TAP task. As it measures direct physical aggression in the particular moment and situation.

These promising results should be interpreted in the context of a number of limitations. First, the lack of a control group and (indirect) measures of interpretation bias that are less closely similar to the training phase means that we cannot completely preclude the possibility that the positive change in interpretation bias is due to some other factors. Future studies should employ measures of interpretation bias that are more different than the training task and/or more indirect in order to be more certain about the impact of the training paradigm on altering interpretations. Second, participants might have been aware of the nature of the experiment, making it possible that demand characteristics played role in the effects of the training. However, if this would truly be an important factor in the current study one might have expected more consistent results across the various measures. Nevertheless, future research could try to include a more unobtrusive training procedure or more unobtrusive outcome measures. Third, future studies with this new paradigm, should encourage transfer of response learning within the study context to participants' perceptions of everyday situations outside the study context. For instance by including more self-relevant processing instructions. Finally, the current study was planned as an initial study and therefore involved a sample of healthy university students. As a consequence it is difficult to make strong inferences about the potential use of the training in a clinical sample.

## Conclusion

The present study provides suggestive evidence that interpretation bias can be modified in a positive direction through the novel CBM-I procedure using visual stimuli, and that this training can have a beneficial effect on mood and self-reported aggressive behavior. The training also seemed to have some effect on a behavioral measure of aggression. These results can be considered an important foundation for further developing and using the current training in research examining the use of CBM-I training as a viable intervention option in treating aggression.

Appendix

Sequence of Wins and Losses of the TAP

| Trial number | Intensity | Duration | Win/Lose |
|---|---|---|---|
| 1 | 0 | 0 | win |
| 2 | 0 | 0 | win |
| 3 | 0 | 0 | win |
| 4 | 0 | 0 | lose |
| 5 | 0 | 0 | lose |
| 6 | 0 | 0 | win |
| 7 | 6 | 7 | lose |
| 8 | 1 | 1 | win |
| 9 | 6 | 5 | lose |
| 10 | 3 | 7 | lose |
| 11 | 5 | 2 | lose |
| 12 | 5 | 9 | win |
| 13 | 2 | 6 | lose |
| 14 | 1 | 3 | win |
| 15 | 3 | 3 | win |
| 16 | 6 | 5 | lose |
| 17 | 10 | 2 | win |
| 18 | 4 | 6 | win |
| 19 | 7 | 9 | lose |
| 20 | 3 | 10 | lose |
| 21 | 6 | 5 | win |
| 22 | 2 | 10 | lose |
| 23 | 10 | 6 | lose |
| 24 | 4 | 10 | win |
| 25 | 9 | 10 | lose |
| 26 | 6 | 4 | win |
| 27 | 2 | 3 | lose |
| 28 | 9 | 7 | lose |
| 29 | 10 | 3 | win |
| 30 | 2 | 6 | lose |

# 3

# CHAPTER

## Gaze-contingent Attention Bias Modification Training and its Effect on Attention, Interpretations, Mood, and Aggressive Behavior

**This chapter has been published as:**
AlMoghrabi, N., Huijding, J., Mayer, B., & Franken, I. H. (2019). Gaze-contingent attention bias modification training and its effect on attention, interpretations, mood, and aggressive behavior. *Cognitive Therapy and Research, 43*, 861-873.

**Abstract**

Cognitive theories propose that aggression is associated with specific patterns of attention to social cues, and suggest that cognitive biases in attention and interpretation are interrelated, The current study tested whether these attention patterns can be altered using a single session of a novel gaze-contingent cognitive bias modification paradigm (CBM-A) and assessed the impact of this on interpretation bias, aggressive behavior and mood. University students (18–31 years) were randomly assigned to either a single session of positive training ($n$ = 40) aimed at increasing attention to pro-social cues, or negative training ($n$ = 40) aimed at increasing attention to negative cues. Results showed that the positive training indeed resulted in an increase in pro-social attention bias, while the negative training seemed not to have an effect on attention to negative cues. Both groups did not differ on their interpretations, mood levels, self-reported aggression and behavioral aggression. Findings suggest that this novel gaze-contingent CBM-A paradigm can indeed alter biased gaze processes, but may not impact interpretations, aggression and mood. The current study was conducted in a non-clinical sample, further research with a clinical aggressive sample, such as forensic patients is necessary to further explore these issues.



**Acknowledgements**

We would like to thank Christiaan Tieman and Marcel Boom from the Erasmus Behavioral Lab for their help in programming the experiment.

The Social Information Processing (SIP) model (Crick & Dodge, 1994) is an influential cognitive theory concerning the development of aggressive behavior. This model asserts that aggressive behavior is associated with specific patterns of social information processing. Several studies that aimed to test this model found support for the existence of these associations suggesting that aggression is associated with biases in both selective attention (e.g., Dodge, 2006) and interpretation of ambiguously hostile behaviors (e.g., de Castro, Veerman, Koops, Joop, & Monshouwer, 2002 for a review). Moreover, different forms of information biases are associated rather than independent phenomena (Crick & Dodge, 1994). Based on the SIP model, it can be hypothesized that reducing aggression-related cognitive biases in attention and interpretation may affect aggression, and furthermore that reductions in one type of bias may affect the other type of bias (c.f. Amir, Bomyea, and Beard, 2010). The ultimate goal of the current study was to test a new attentional bias modification training and assess its effects on attention, interpretations, mood and aggressive behavior. A logical starting point of this endeavor is focusing on how aggressive individuals differ in their attentional deployment from non-aggressive individuals.

According to the SIP model (Crick & Dodge, 1994), individuals first attend to the most relevant social cues in a social situation and encode it for further processing. Encoding functions in a bottom-up manner that affects the way the social situation is interpreted. Thus, encoding has to be selective and fast in order to efficiently identify all relevant cues in the environment. The traditional hypothesis of the SIP model suggests that aggressive individuals tend to show heightened attention for hostile versus non-hostile social cues, increasing the likelihood of a hostile interpretation of the situation, therefore increasing the chances of aggression (Crick & Dodge, 1994). In support of this hypothesis a number of studies found that individuals who score high on measures of aggression or anger tend to show heightened attention for hostile stimuli on various reaction-based tasks, like the dot-probe (e.g., Smith & Waterman, 2003, but see Schippell, Vasey, Cravens-Brown, and Bretveld, 2003), the emotional Stroop (e.g., Eckhardt & Cohen, 1997; Smith & Waterman, 2003; Van Honk et al., 2001a; Van Honk et al., 2001b), and visual search (e.g., Cohen, Eckhardt, & Schagat, 1998; Smith & Waterman, 2004). However, almost all these studies used verbal stimuli (but see Van Honk et al., 2001a) that were presented without a context. As a result the patterns of attentional deployment captured by such paradigms may not be optimally informative of attentional processes during actual social interactions.

To overcome such issues, other studies have focused on attention deployment to visual stimuli depicting social situations, using eye-tracking (Wilkowski, Robinson, Gordon, & Troop-Gordon, 2007; Horsley, de Castro, & van der Schoot, 2010; Troop-Gordon, Gordon, Vogel-Ciernia, Lee, & Visconti, 2018). Interestingly, these studies show a different pattern of results, supporting an alternative hypothesis described as the 'schema inconsistency hypothesis'. According to this hypothesis aggressive individuals' interpretations of social situations are based more on pre-existing hostile intent schemata than on available social cues in the current social situation. Importantly, even though some studies suggest that aggressive individuals focus their attention on schema inconsistent cues (i.e., non-hostile cues) (Wilkowski et al., 2007, Horsley et al., 2010), these cues are not well recalled (Horseley et al., 2010) suggesting that schema-inconsistent information is sub-optimally encoded (de Castro & van Dijk, 2017). In order to test this idea, Troop-Gordon and colleagues (2018) presented children with video clips of child actors portraying scenes of ambiguous provocation, and assessed their peer beliefs. They found that aggressive children who hold negative peer beliefs take greater time before they first fixate on social cues from the actors in the scene, in particular the provocateur, while they do not dwell longer on the provocateur after the actual provocation has occurred. Such initial inattention to social cues, and the failure to compensate for this after a provocation, may be a result of overreliance on schema-based hostile beliefs in the context of ambiguous situations. Taken together, the findings from these studies suggest that aggressive individuals might benefit most from training programs that would train them to effectively attend to and encode relevant social cues that help disambiguate the situation. Therefore, the current study assessed the effect of an attention training program aimed at explicitly directing attention towards relevant social cues while trying to determine the intent of an actor in ambiguous social situations.

One way to train attention, is to use the CBM-A paradigm. CBM-A was originally developed to manipulate attention selectivity in the context of anxiety research where it is used to change participant's attention selectivity away from threatening cues to more non-threatening cues (MacLeod, Rutherford, Campbell, Ebsworthy, & Holker, 2002). Studies have shown that such manipulations of attentional bias influenced anxiety and stress reactivity (see Bar-Haim, 2010 for review). However, the results have been mixed and the reported effect sizes are small to moderate (Van Bockstaele et al., 2013). This may have to do with the fact that CBM-A procedures that have been used so far inferred focus of attention on the basis of manual reaction times to visual cues on

the screen. This makes it difficult to ascertain whether the training indeed affects visual direction of attention. A more powerful and direct manipulation would be to provide feedback based directly on the gaze direction using an eye tracker. Therefore, the current study used a novel gaze-contingent CBM-A procedure, which potentially has better effects in training attention in the context of aggression.

Recent studies in the context of depression and anxiety show that attention can indeed be trained successfully using gaze-contingencies (Price, Greven, Siegle, Koster, & De Raedt, 2016; Ferrari, Mobius, van Opdorp, Becker, & Rinck, 2016, Lazarov, Pine, & Bar-Haim, 2017). Following this, in the present training, a gaze-contingent procedure in which the screen is updated based on the individual's eye position (Foulsham, Gray, Nasiopoulos, & Kingstone, 2013), was used to manipulate attention. More specifically, we provided positive feedback to participants if they fixed their gaze on the pro-social cues, and negative feedback if they fixed their gaze on the negative cues in ambiguous social provocation scenes. Such a setup might potentially increase the training effects as it ensures a fixation on and processing of the information in the desired areas of interest. Importantly, it provides an effective real time attention manipulation of the cues (Glaholt & Reingold, 2011).

In the current study, the CBM-A training provided a first step toward the development of attention bias training program aimed at training more pro-social looking strategies for aggressive individuals. During the training participants were presented with pictures of ambiguous social situations in which something unfortunate happens (e.g., one person spilling a drink on someone else). Previously it has been shown that individuals scoring high on aggressive tendencies tend to pay less attention to the face of a potential harm-doer (i.e., provocateur) in scenes depicting ambiguous signs of hostility, and tend to look longer at angry body expressions, than do individuals scoring low on aggressive tendencies (Lin et al., 2016). Arguably, the face is the single most informative social cue regarding the intentions of one person towards another (Cadesky, Mota, & Schachar, 2000). Following this, directing individual's attention to facial expressions during social interactions may provide a viable target in CBM-A training. In addition, by combining the attention training with the explicit instruction to look at cues that can help disambiguate the situation, we hoped to ensure encoding of the attended information. In the current CBM-A two cues were identified on each picture; pro-social cues which includes the face of the harm-doer, which can indicate whether the incident happened by accident (or not); or to negative cues (e.g.,

the drink spilling on victim) which provides no useful information regarding the intent of the harm-doer and might only increase feelings of anger in the participant. Depending on the training condition, participants were either trained to attend more to the pro-social cues or to the negative cues.

The current study had four aims. First, we aimed to examine whether aggression-related attention mechanisms can be altered using this novel gaze-contingent CBM-A procedure. Second, we aimed to examine the effects of the altered aggression-related attention mechanisms on aggressive behavior using self-report and behavioral measures. We predicted that training individuals to attend to the negative cues would increase subsequent attention bias to negative cues and increase aggressive behavior. On the other hand, training them to attend to the pro-social cues would increase pro-social attention and reduce subsequent aggressive behavior. Third, this study aimed to test whether this procedure affects how subsequent ambiguous social information is interpreted, in order to investigate the interaction between attention and interpretation bias and how both of these biases contribute to aggressive behavior. This is relevant because it can show whether CBM procedures need to target only one or better target both biases to achieve the strongest effects. We expected that participants who were trained to attend to pro-social cues would make less hostile interpretations than participants who were trained to attend to negative cues. Finally, based on previous research in the context of anxiety (MacLeod et al., 2002) showing that manipulating attention bias may impact mood, we also assessed the impact of the attention modification training on mood in an explorative way.

## Method

### Participants

Forty male and forty female students from Erasmus University Rotterdam (48 Caucasians, 5 Asian, 7 Middle Eastern, 2 Hispanic, 1 African, and 17 others), aged between 18 and 31 ($M$ = 20.61, $SD$ = 2.11) participated in exchange for course credits. Participants were randomly selected from a list of students who had subscribed to participate in the experiment. The study was conducted according to the rules of the Helsinki Declaration on informed consent and confidentiality (World Medical Association, 2001) and all procedures were carried out with adequate understanding and written consent of the participants.

**Eye-tracking procedure**

During the CBM-A training, eye movements were recorded using a SMI-RED 250 device (Sensomotoric Instruments GmbH, Teltow, Germany) with a sampling rate of 250 Hz.

The stimuli were presented on a 22-inch computer screen with a resolution of 1,680 x 1,050 pixels. The viewing distance was approximately 60 cm. The size of the picture was 1,344 x 777 pixels. For each image, areas of interest (AOI) were defined around a 'negative' cue showing the negative outcome of the situation (e.g., coffee spilling on the victims clothes), and a 'pro-social' cue (the face of the harm-doer, see Figure 2). Each AOI was defined as a square area and had a size of either 252 x 210 or 336 x 210 pixels to encompass the entire area of display of pro-social or negative cue in the picture.

To ensure accuracy of the gaze pattern, a nine-point calibration and 4-point validation was performed before starting with the first phase. Also, a chin-rest was used to maintain a constant head position and distance from the computer screen throughout the training.

**CBM-A Training**

The CBM-A task consisted of 52 trials that were presented using E-prime software. On each trial, the participants viewed an image of a social interaction during which something unfortunate happens, like one person spilling a drink over the other, while the intention of the harm-doer is unclear. These images were used to assess attention and interpretation biases and manipulate attention bias. Each image appeared only once, so 52 different pictures were used. The training task was completed within a single session and started with an eye-tracker calibration. The CBM-A training consisted of four phases: practice, baseline, training, and test. The practice phase was implemented to introduce participants to the experimental procedure and consisted of three trials. In order to examine the effects of the training on attention and interpretation bias, an assessment of attention and interpretation bias was administered during the baseline and test phases. The baseline and test phases were identical and consisted of six trials each. The manipulation of attention bias took place during the training phase, which consisted of forty trials. The whole CBM-A task took approximately 25 min to complete.

**Phase 1 (practice).** On each trial participants were presented with an image which is not related to the images used in the training. To get acquainted with the procedure, participants were instructed to fix their gaze on a certain AOI and received feedback on the their performance; "Correct" if they fixed their gaze on the correct part of the picture; "Incorrect" if they fixed their

gaze on the incorrect part of the picture; or "Too slow" if they didn't fix their gaze on any AOI and were asked to try again.

**Phase 2 (baseline) and 4 (test).** On each trial participants were presented on the computer screen with a single sentence describing a situation in which a mishap has occurred. For example, "There is water all over his clothes!" The description was presented on the screen until the mouse was clicked. Participants were then presented with an image of the described situation in which the intent of the harm-doer was ambiguous (see Figure 1 for an example). While looking at the images, participant's eye movements were recorded automatically using the eye-tracking device. During these phases participants' total dwell-time to both areas of interest (i.e., pro-social and negative cues) was recorded, which we used as a measure of the attention bias.

To measure attention bias, participants were asked to look at the part of the picture that best indicates whether or not the incident happened on purpose (e.g., see Figure 1). To assess participants' interpretation of the intent of the harm-doer they were asked "Why did this happen?", and presented with two possible interpretations, one hostile and one benign (cf. AlMoghrabi, Huijding, & Franken, 2018). For example, the picture presented in Figure 1 was accompanied by the following two interpretations: (a) This happened on purpose because he wanted to tease him (hostile interpretation); (b) This happened by accident because he tripped (non-hostile interpretation). Participants indicated the likelihood that a specific interpretation is true by dragging an arrow on a 100-point visual analogue scale that was anchored with the labels "No, definitely not" (-50) on the left and "Yes, definitely" (+50) on the right ends of the scale. During this phase, no feedback was provided. The viewing time was fixed for 5000 ms for each image. Additionally, a minimum amount of eye gaze time of 80 ms at a certain AOI was qualified as a gaze fixation (e.g., Huijding, Mayer, Koster, & Muris, 2011; Gerdes, Alpers, & Pauli, 2008).

**Phase 3 (training phase).** For the training phase, the participants were randomly assigned to either the negative or positive training, each consisting of forty trials. Similar to phases 2 and 4, each trial presented participants with an image of a situation in which one person is harming another, but the intention of the harm-doer is unclear. The images were always preceded by a short description of the situation that was presented for 3000 ms. For example, the image presented in Figure 2 was preceded by the description: "He got the ball hard on his head!" Subsequently, the image of the situation was presented on the screen for 5000 ms, along with the question "Why did this happen?" Participants were instructed to fixate on the part of the picture that best indicates

whether the incident happened on purpose or by accident, until they received feedback. In this phase a gaze-contingent procedure was used to ensure participant's fixation on the specified areas of interest. Depending on the training condition either the negative or the pro-social cue was reinforced as the correct answer. In the positive training condition, fixations on the pro-social cues (the faces of the harm-doers) were reinforced as "correct" while in the negative training fixations on the negative cues (the negative outcomes) were reinforced as "correct". If participants fix their gaze for 1000 ms on the "correct" AOI, the word "CORRECT" was presented at the top of the screen in bold green font. If participants fix their gaze for 1000 ms on the "incorrect" AOI, the word "INCORRECT" was presented at the top of the screen in bold red font. This feedback remained on the screen for 2,000 ms, after which the next trial began. If participants didn't fix their gaze on either AOI for 5000 ms "Too slow" was presented on top of the screen in bold blue font for 2000 ms, after which the same picture would be shown to allow the participant to try again.

**Stimulus materials**

A set of 52 pictures was used in the CBM-A training that each showed a situation in which one person harmed another, but was ambiguous regarding the intent of the harm-doer. For the baseline and test phases, we used the images from the study of Wilkowski et al. (2007) (see Figure 1 for an example). For the training phase, we used the images from the study of Horsley et al. (2010) (see Figure 2 for an example), supplemented by thirty images from stock image websites. Images were chosen that depicted a hypothetical real-life scenario, some including two men, some two women, and some a man and a women. The images depicted an interaction between those two characters, with one of the two characters (i.e., harm-doer) initiating a behavior that affects negatively the other character (i.e., victim).

To ensure the adequacy of the stimulus materials, in a pilot-study 40 university students were asked to rate the pictures on a number of characteristics, including the extent to which the depicted harm was intentional and how aggressive is the facial expression of the harm-doer. Participants rated intentionality on a 100 point visual analogue scale (VAS) that was anchored with the labels "Accidental" on the left and "Intentional" on the right end. Additionally, participants rated the facial expression of the harm-doer on a 100 point VAS that was anchored with the labels "Friendly" on the left and "Aggressive" on the right end. The results show that the pictures in the assessment phase were rated on average as very ambiguous regarding both the intent of the harm-doer [$M$ = 51.3, $SD$ = 14.1], and facial expression of the harm-doer [$M$ = 50.8, $SD$ = 6.5], and the

pictures in the training phase were rated ambiguous regarding the intent of the harm-doer [*M* = 47.0, *SD* = 11.6], and quite ambiguous, but leaning a bit towards friendly, for the facial expressions of the harm-doer [*M* = 41.76, *SD* = 4.8].

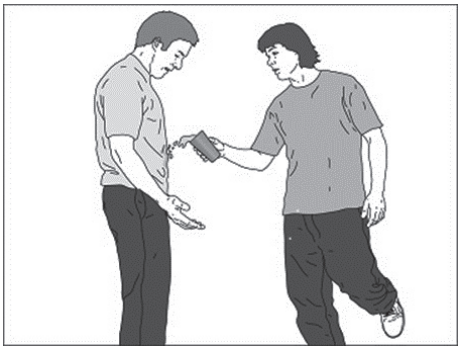


*Figure 1*. Example image from the baseline phase.

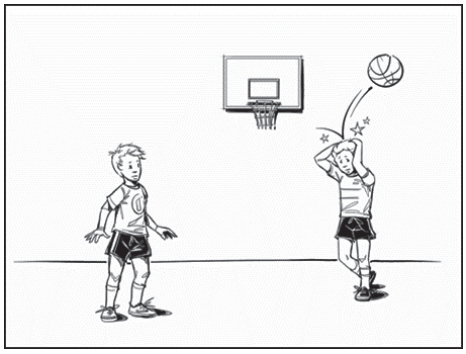


*Figure 2*. Example image from the training phase.

**Pre-measures**

Prior to the CBM-A training, the present study sought to assess participants on a number of measures of state/trait aggression, anxiety, mood, and anger.

The Buss and Perry's (1992) trait Aggression Questionnaire (AQ) assesses trait aggression. Following the same method used by Farrar and Krcmar (2006), the present study reworded the AQ measure to assess state aggression (cf. AlMoghrabi et al., 2018). The modified questionnaire started with the following instruction: "Imagine that you just bought something to drink. When

you walk outside, somebody bumps into you, spilling your drink over your favorite clothes. As you look at the mess, you hear this person swearing." In addition, the items comprised of items from the AQ that were rephrased. For example, the original AQ item "I have trouble controlling my temper" was rephrased to "I would have trouble controlling my temper with this person" to match state aggression. For each of the items, the participants were instructed to rate the extent (1 = extremely uncharacteristic of me; 7 = extremely characteristic of me). The questionnaire consists of 20-items on three subscales: physical aggression, verbal aggression, and anger. In the current sample, Cronbach's alpha was .87.

The Reactive-Proactive Aggression Questionnaire (RPQ; Raine et al., 2006) provides measures of both reactive (11 items; e.g., "damaged things because you felt mad") and proactive (12 items; e.g., "taken things from other students") aggression. For each item the participant provided a rating of 0 = Never, 1 = Sometimes, and 2 = Often. In the current sample, Cronbach's alpha was .77 for reactive and .75 for proactive aggression. Finally, anger was measured using part B of the Novaco Anger Scale (NAS; Novaco, 1994). The measure consists of 25 potentially provoking situations (e.g., "Being joked about or teased"). The participant rated each provoking situation on a 5-point scale from 0 (little or no annoyance) to 4 (very angry). In the current sample, Cronbach's alpha was .88. Additionally, the participant's state mood was measured pre-training by asking participants to rate how happy, angry, sad, and afraid they felt at the moment. For each emotion they dragged an arrow on a 100-point visual analogue scale that was anchored with the labels "Not at all" (-50) on the left and "Very much" (+50) at the extreme ends of the scale.

For exploratory purposes beyond the scope of this manuscript the State-Trait Anxiety Inventory (STAI) was also included (Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983).

**Post-measures**

To test whether the training would influence self-reported aggression, the participants completed post-training again the reworded trait Aggression Questionnaire but with a different contextual story that read: "Imagine that you are at the Starbucks working on an assignment. Suddenly, someone bumps into your table, spilling coffee all over your notes. You see that the other person looks really annoyed." In our sample, Cronbach's alpha was .89.

The Positive Affect and Negative Affect Schedule (PANAS; Watson et al., 1988) was administered post-training to measure trait mood levels. Participants had to rate how much they generally feel (1 = Slightly; 5 = Extremely) about 10 positive emotional states (e.g., interested, inspired) and 15 negative states (5 items specifically covering anger were added to the original; e.g., upset, guilty). Cronbach's alpha for positive effects was .87, and for negative effects was .92. Additionally, the participant's state mood was measured again post-training by asking participants to rate how happy, angry, sad, and afraid they felt at the moment. For each emotion they dragged an arrow on a 100-point visual analogue scale that was anchored with the labels "Not at all" (-50) on the left and "Very much" (+50) at the extreme ends of the scale.

**Aggression Task**

In addition to the self-reported measures, aggression was also measured post-training using the Taylor Aggression Paradigm (TAP; Taylor, 1967) which is a behavioral measure of aggression. The task was introduced to the participants as a competitive reaction time game of 30 trials, and they were told that they would be competing against an opponent. Before starting with the actual task, the experimenter gave a brief introduction by telling each participant that this experiment was a collaboration between Erasmus University Rotterdam and Utrecht University and that their opponent was currently present at a lab in Utrecht and that the same instructions would be delivered to their opponent. After this, the experimenter would pretended to contact collaborators at Utrecht University to coordinate the start time of the experiment. This was done to ensure the credibility of the game. In fact, no experimental collaboration or opponent actually existed.

Each participant was seated at a desk with a mouse and a computer screen, and told that in order to beat their opponent in this reaction time game, they had to click the mouse as fast as possible when a rectangle turned from yellow to red. Participants were instructed that if they received the message "You Won" it would mean that they clicked faster than their opponent, while the message "You Lost" meant they were slower. Participants were informed that the winner would be allowed to administer a noise blast to their opponent. To make it more believable, the game started with the message "Connecting with opponent" on the screen. Also, in order to give the participant an idea of what kind of noise stimulus was used in the task in terms of intensity and duration, a noise testing procedure was administered before commencing the real task. Following that, on each trial participants first selected the duration (between 0 and 10 seconds) and the

volume of the noise blast (between 0 and 100 dB) they would administer to the opponent should they win the trial. When they "lost" a trial, participants received a noise blast through the headphones and were given feedback regarding the level and duration of the noise they had received from their opponent. When participants "won" a trial, they could see on the screen what duration and level of noise their opponent's had set at the beginning of the trial. The opponent's noise selections, as well as the order of winning and losing trials, was pre-programmed (for the sequence of wins and losses; cf. Brugman et al., 2015).

**Procedure**

The participants were quasi-randomly assigned to one of two conditions: the positive condition ($n$ = 40; 20 males and 20 females), which aimed to increase attention bias to pro-social cues or the negative condition ($n$ = 40; 20 males and 20 females), which aimed to increase attention bias to negative cues. For either condition, the experimenter would start with a short introduction and a general explanation of the experimental tasks. Following this, participants started by completing the AQ, STAI, RPQ, and NAS questionnaires. Subsequently, they received specific instructions regarding the eye-tracking and the CBM-A training. After completing the CBM-A training the experimenter explained the TAP. After making sure that the participants understood the instructions of the TAP, they then proceeded with the task. Finally, the participants completed the AQ and PANAS. The entire experiment took approximately 60 min to complete.

## Results

**Data reduction and preliminary analysis**

First, based on the eye-tracking data, we calculated separate mean total viewing times in ms for the pre-defined AOIs for the pro-social and the negative cues at pre- and post-training. Next, pre- and post-training attention bias (AB) scores were calculated by subtracting the mean total viewing time at the negative cues from the mean total viewing time at the pro-social cues. Thus, a higher AB score indicates more attention allocation to pro-social (facial) than to negative (negative outcome) cues. Also, we calculated separate interpretation bias (IB) scores for each condition for the pre- and post-training assessments by subtracting the mean VAS likelihood rating for the hostile interpretation to be true from the mean VAS likelihood rating for the pro-social

interpretation to be true. Thus, positive IB scores indicate that pro-social interpretations were rated as more likely to be true than hostile interpretations.

Next, in order to ascertain the appropriateness of our AB measure, we correlated the attention bias scores (AB-pre and AB-post) with the concurrently assessed aggression-related measures (i.e., AQ, NAS, RPQ, TAP and VAS state anger). The results indicated that there were no significant relations between pre- and post-training attention bias scores with respectively pre- and post-training aggression-related measures (see Table 1).

Table 1

*Correlations between Attention Bias Scores Pre/Post-Training and Aggression-Related Measures Pre/Post-Training*

| Measures | Attention bias |
|---|---|
| Aggression Questionnaire | -.54 / -.06 |
| Physical Aggression | -.05 / -.12 |
| Verbal Aggression | -.02 / .06 |
| Anger | -.05 / -.08 |
| Reactive-Proactive Aggression Questionnaire | .01 / n.a. |
| NAS | -.08 / n.a. |
| PANAS-positive | n.a. / .05 |
| PANAS-negative | n.a. / -.13 |
| Angry mood | .06 / -.21 |
| Afraid mood | -.02 / -.13 |
| Sad mood | -.07 / -.09 |
| Happy mood | .01 / -.07 |
| Taylor Aggression Paradigm | n.a. / -.08 |
| Intensity | n.a. / -.08 |
| Duration | n.a. / -.13 |

*Note.* n.a. = not assessed; NAS = Novaco Anger Scale; PANAS = Positive Affect and Negative Affect Schedule.

All correlations: $p > .05$.

**Baseline measures**

There were no significant differences between the participants in the positive and negative training conditions in their baseline levels of self-reported aggressive behavior (AQ and RPQ), anger (NAS), trait anxiety (STAI-T), and mood ratings (happy, angry, sad, and afraid), for all $t(78) < -1.16$, $p > .201$. However, participants in the positive training condition reported a higher level of pre-training state anxiety (STAI-S) than participants in the negative training condition, $t(78) =$

2.39, $p = .019$. Descriptive statistics for the pre- and post-training measures are presented in Table 2. In addition, the analysis showed that participants in the negative groups scored higher on pro-social interpretation bias prior to the training ($M = 9.53$, $SD = 23.00$) than participants in the positive group ($M = -1.39$, $SD = 22.63$), $t(78) = -2.14$, $p = .035$. Both groups did not differ significantly on attention bias prior to the training $t(78) = 1.50$, $p = .137$, for the negative group ($M = -610.85$, $SD = 1458.02$) and for the positive group ($M = -167.70$, $SD = 1165.56$).

Table 2

*Descriptive Statistics for Pre/Post-Training Measures*

| Measures | Positive training | | Negative training | |
|---|---|---|---|---|
| | *M* | *SD* | *M* | *SD* |
| Pre-training | | | | |
| Aggression Questionnaire | 65.55 | 17.39 | 64.88 | 14.85 |
| Physical Aggression | 23.93 | 8.78 | 24.98 | 8.05 |
| Verbal Aggression | 18.33 | 4.86 | 17.70 | 4.16 |
| Anger | 23.30 | 6.78 | 22.20 | 5.91 |
| Reactive-Proactive Aggression Questionnaire | 32.18 | 5.06 | 32.33 | 6.10 |
| NAS | 71.85 | 13.58 | 70.85 | 12.52 |
| Anxiety Inventory-State | 36.70 | 10.80 | 31.95 | 6.48 |
| Anxiety Inventory-Trait | 42.55 | 10.06 | 42.43 | 7.84 |
| Angry mood | -40.98 | 16.78 | -41.70 | 12.76 |
| Afraid mood | -36.05 | 23.81 | -41.53 | 12.47 |
| Sad mood | -32.78 | 23.97 | -36.73 | 16.70 |
| Happy mood | 13.35 | 19.51 | 18.20 | 18.00 |
| Post-training | | | | |
| Aggression Questionnaire | 64.45 | 18.98 | 63.15 | 15.81 |
| Physical Aggression | 24.55 | 8.81 | 24.68 | 8.08 |
| Verbal Aggression | 18.23 | 5.85 | 17.22 | 5.29 |
| Anger | 21.67 | 7.42 | 21.25 | 5.86 |
| PANAS-positive | 27.45 | 7.79 | 26.35 | 5.86 |
| PANAS-negative | 22.45 | 9.10 | 21.33 | 6.37 |
| Angry mood | -35.28 | 22.41 | -35.38 | 20.35 |
| Afraid mood | -37.35 | 21.22 | -42.83 | 13.67 |
| Sad mood | -31.73 | 19.89 | -35.23 | 18.39 |
| Happy mood | 12.03 | 22.04 | 16.55 | 19.84 |
| Taylor Aggression Paradigm | 16.78 | 12.20 | 19.62 | 15.64 |

*Note.* NAS = Novaco Anger Scale; PANAS = Positive Affect and Negative Affect Schedule.

**Reliability of the attentional process measures**

To assess the reliability of the attentional bias measure Cronbach's alpha's were calculated separately for baseline and test phase. First, we calculated separate total viewing times in ms for the pre-defined AOIs for the pro-social and the negative cues at pre- and post-training. Trials with less than 80 ms at either areas of interest were excluded. From the whole sample one participant looked less than 80 ms at either areas of interest on one trial. As a result we were unable to take this trial into account. Next, pre- and post-training attention bias scores for each image were calculated separately by subtracting the total viewing time of the negative cues from the total viewing time of the pro-social cues. The Cronbach's alpha values for the pre- and post-training bias scores in the current sample were (baseline phase: $\alpha = .86$; test phase $\alpha = .84$).

**Effects of attention training on attention bias**

To determine training effects on attention bias, AB scores were subjected to a 2 Assessment (pre, post-treatment) x 2 Group (positive versus negative training) ANOVA with repeated measures.

The analysis revealed significant main effects of Group, $F(1, 78) = 21.43$, $p < .001$, $\eta_p^2 = .22$, and Assessment, $F(1,78) = 8.58$, $p < .01$, $\eta_p^2 = .10$. More importantly, the crucial interaction between Group and Assessment was significant, $F(1, 78) = 15.04$, $p < .001$, $\eta_p^2 = .16$ (see Figure 3). This interaction was decomposed using paired-samples *t*-tests of change over time. This showed that in the positive condition, attention bias became significantly more positive, indicated relatively longer fixation durations on the pro-social cues (i.e., the face of the harm-doer) then on the negative cues: $t(39) = -5.43$ $p < .001$. In the negative condition, attention bias scores did not change significantly over time: $t(39) = .61$, $p = .546$.

Inspection of the participants' accuracy during the training phase (i.e., the extent to which they were doing what we wanted them to do during the training) showed that participants in the negative training condition made significantly fewer errors ($M = 17.56\%$, $SD = 11.26$) as compared to participants in the positive condition ($M = 24.94\%$, $SD = 20.06$, $t(78) = -2.03$, $p < .05$). This suggests that the observed difference in training effects between the two conditions cannot simply be attributed to differences in compliance to the training instructions. That is, compliance to the training instructions was significantly greater in the negative than in the positive condition, while the effects of the training on attention were greater in the positive than in the negative condition.

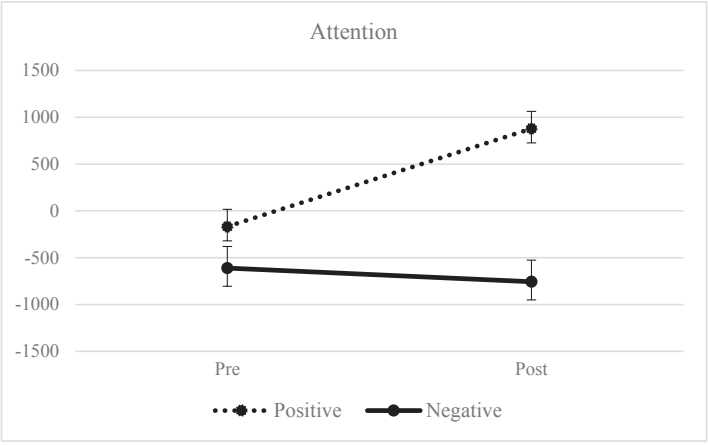



*Figure 3*. Average attentional bias scores at pre- and post-training for each training condition. Error bars indicate standard error of the mean.

**Effects of attention training on interpretation bias**

To examine the effects of the attention training on interpretation bias, the IB scores were subjected to a 2 Assessment (pre versus post-treatment) x 2 Group (negative versus positive training) ANOVA with repeated measures. The analysis revealed that the crucial interaction between Group and Assessment was not significant, $F(1, 78) = 1.50$, $p = .224$, $\eta_p^2 = .02$. Moreover, no significant effects for Group emerged, $F(1,78) = 2.43$, $p = .123$, $\eta_p^2 = .03$. However, the main effect of Assessment was significant, $F(1, 78) = 62.97$, $p < .001$, $\eta_p^2 = .45$. Surprisingly, it was found that in both conditions interpretation bias became significantly more pro-social post training (see Figure 4).

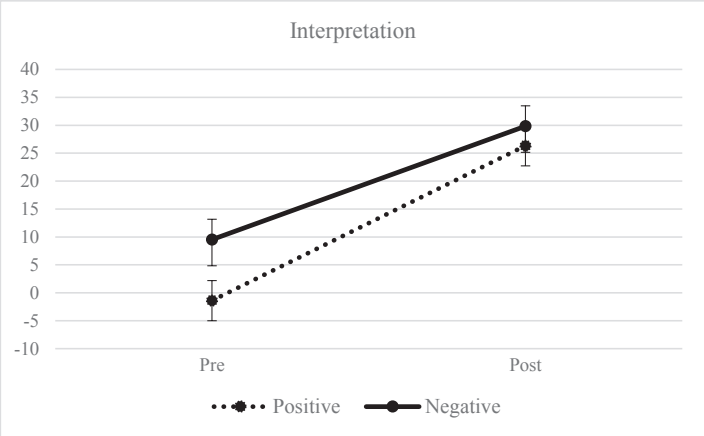



*Figure 4*. Average interpretation bias scores at pre- and post-training for each training condition. Error bars indicate standard error of the mean.

**Effects of attention training on mood**

VAS state mood ratings (happy, angry, sad, and afraid) were subjected to separate 2 Assessment (pre versus post-treatment) x 2 Group (positive versus negative training) ANOVAs with repeated measures. Only a significant main effect of Assessment emerged for self-reported anger, $F(1, 78) = 7.76$, $p < .01$, $\eta_p^2 = .09$, indicating that in both conditions self-reported state anger significantly increased from pre- to post-training. None of the other effects were significant, for all $F(1,78) < .02$, $p > .885$, $\eta_p^2 = .00$.

In addition, independent-samples *t*-tests on the PANAS scores confirmed that the positive and the negative condition didn't differ significantly in terms of either their positive or negative trait affect scores, for both $t(78) < .64$, $p > .477$.

**Effects of attention training on aggression**

Participants scores from the AQ were subjected to a 2 Assessment (pre- versus post-treatment) x 2 Group (positive versus negative training) ANOVA with repeated measures. The analysis revealed no main effects of Group or Assessment and no significant interaction between Group and Assessment, $F(1, 78) = .08$, $p = .774$, $\eta_p^2 = .00$ (see Figure 5). Additionally, the analysis revealed that the training did not result in changes on the AQ subscales, all $F(1.78) < .66$, $p > .421$, $\eta_p^2 > .003$.

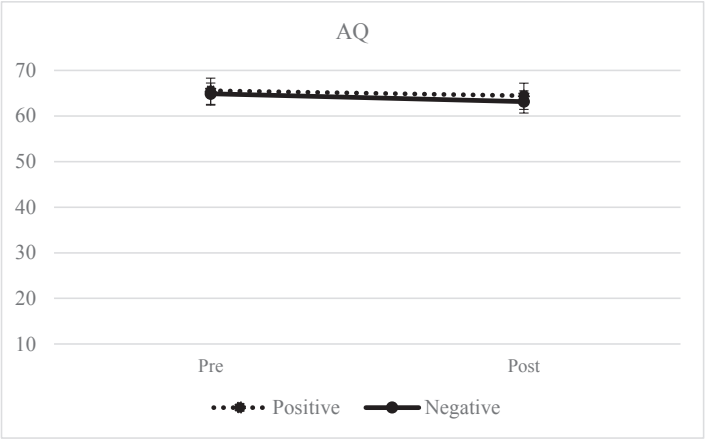



*Figure 5*. Average Aggression Questionnaire (AQ) ratings at pre- and post-training for each training condition. Error bars indicate standard error of the mean.

Finally, participant's TAP scores were compared between the two conditions. An independent-samples *t*-test showed that the two training groups did not differ in terms of their TAP performance ($t(78) = -0.91$, $p = .367$), intensity ($t(78) = -.20$, $p = .845$), and duration ($t(78) = -.97$, $p = .337$).

## Discussion

The current study examined whether a novel gaze-contingent cognitive bias modification of attention (CBM-A) procedure -designed to modify attention bias using pictorial stimuli- influences attention, interpretations, mood and aggressive behavior. Results indicate that gaze-contingent attention training within the positive condition indeed resulted in an increase in attention to pro-social (facial) cues in images of ambiguous social situations. However, no change in attention to either pro-social or negative cues was found in the negative condition. Moreover, the attentional bias scores were unrelated to the concurrently assessed aggression related measures. Additionally, in both the positive and negative attention training conditions interpretations changed in a pro-social direction, and increased self-reported state anger was found.

The current finding that the positive training increased pro-social attention bias is well in line with previous findings that attention bias can be trained (Amir, Beard, Burns & Bomyea 2009; Amir et al., 2009; Van Bockstaele et al., 2013; Wadlinger & Isaacowitz, 2008). Moreover, this finding underscores the feasibility of using a gaze-contingent approach to training attentional

deployment (Price et al., 2016; Ferrari et al., 2016; Lazarov et al., 2017). The gaze-contingent approach was successful in training participants in the positive condition to pay more attention to pro-social cues (i.e., the face of the harm-doer) than to negative cues (i.e., the negative outcome) in a picture of an ambiguous social situation. The major advantage of this procedure is that the set-up enables direct assessment and training of gaze direction, rather than inferring this on the basis of task performance (i.e., reaction times) as is usually the case in attentional bias modification procedures. In addition, the current approach allows participants to experience the effect of their own eye-movements on altering the on-screen view presented to them, which creates interactive and responsive stimuli.

In contrast, although it appears that in the negative condition there was a slight increase in viewing negative cues from pre- to post-training, the attentional bias change score for this condition was not significant. This lack of training effect might be related to the fact that at pre-training, participants in both groups spent more time looking at the negative cues than the pro-social cues, suggesting that the negative cues were most salient in the depicted social situations. This is in line with a study of Wadlinger and Isaacowitz (2008) that found that participants looked longer at negative stimuli post neutral attention training, and argued that if participants were not trained to attend less to negative cues, these cues may be considered as "attention grabbing" in a social situation. Similarly, Ferrari and colleagues (2016), who also used a gaze-contingent attention bias modification procedures in a healthy sample, found that at pre-training participants took longer to disengage from negative stimuli than from positive stimuli. They argued that it takes more time to disengage from high arousing stimuli which in this case were the negative or threat-related stimuli. This might explain why our current sample in both conditions didn't show pro-social attentional bias pre-training which is supposed to be typical for healthy individuals. Additionally, the pre-existing negative attentional bias in the negative condition might also explain why participants in this condition have made very few errors in the training phase. That is, the training was reinforcing this pattern of selective attention toward negative cues, resulting in no further significant increase in negative attention bias pre- to post-training.

In general, the attention training did not have any effect on the aggression measures post-training. Additionally, the results showed that the attention training did not appear to have an effect on the TAP as a behavioral measure of aggression. Likewise, the attention bias scores did not correlate with the TAP scores and self-reported aggression scores both pre- and post-training.

Consequently, the current study was not able to provide evidence for the association between attention bias and aggression. Furthermore, the CBM-A training did not result in the expected effects on interpretation bias. Earlier we argued that the face may be the single most informative social cue regarding the intentions of one person towards another (Cadesky et al., 2000). Therefore, it was suggested that high trait angry and aggressive individuals may have trouble mitigating their initial hostile interpretations, because they do not pay enough attention to and/or may not encode the right social cues. Following this line of reasoning, we hypothesized that aggressive individuals might benefit from training programs that would help them to effectively attend to relevant social cues that will help disambiguate the environment. Our current results suggest that this is not the case. That is, we did not find differential effects of training participants to attend to the pro-social (facial) cues or negative (outcome) cues on participants' interpretations of the ambiguous situations. In prior anxiety research, it has been indicated that cognitive biases influence and interact with one another in maintaining social anxiety (Amir et al., 2010; Hirsch & Clark, 2004; Hirsch, Clark, & Mathews, 2006; White, Suway, Pine, Bar-Haim, & Fox, 2011). For example, in the study of White et al. (2011) participants who were trained to attend to threat cues were more likely to interpret ambiguous stimuli as threat-related as compared to participants in a placebo-training group. Also, Amir et al. (2010) provided evidence that a single session of interpretation modification program modified interpretation bias in social anxiety participants, which in turn led to an increase in their ability to disengage attention from threat stimuli. Despite the hypothesis that modification of attention bias may influence interpretation bias, vice versa and thus enable changes on aggression, focusing on one cognitive bias may be insufficient to cause change in the context of aggression. In this case future work on CBM should target more biases at the same time, which may enable stronger training structure for it to become more malleable. In another line of argument, it could be that the current findings fit better with the reasoning of Wilkowski et al. 2007 and Horsley et al. 2010, that interventions targeting attention allocation should not only target attention allocation toward mitigating cues but to also target schemas that triggers a hostile interpretation of encoded cues in a social situation. In support of this idea, in a previous study using a similar training that was aimed at retraining hostile interpretations we did find some effect on aggressive outcomes (AlMoghrabi et al., 2018).

Additionally, the current CBM-A training did not result in expected effects on state mood, since self-reported angry mood state had increased in both conditions. This fits best with the

findings of Ferrari et al. (2016) who found that negative mood increased in both negative and positive training groups post a gaze-contingent attention training. However, the negative group showed a stronger increase in negative mood than the positive group. This suggests that the increase in negative mood in the negative group might be due to sustained attentional processing of negative stimuli. In our case, the increase in self-reported angry mood from pre- to post-training in the negative condition could be related to the fact that participants had to continuously attend and process negative social cues during the training and were reinforced for a correct response. While in the positive condition the increase in self-reported angry mood from pre- to post-training might be the result of the high number of errors that participants made during the training compared to the participants in the negative condition. It is possible that participants in the positive training were inclined to fix their attention on negative cues when their attention should be fixed on pro-social cues, and became angry or annoyed by repeatedly receiving negative feedback. However, it is important to note that our sample did not include aggressive or high trait angry participants, making it more difficult to find aggression-related effects. Future research could apply this training to a clinically aggressive sample, before drawing firm conclusions about its therapeutic value.

The current results should be taken in light of several limitations. First, the current study included a sample of healthy university students. Therefore, it is not possible to make strong inferences about the potential use of the training in a clinical sample of aggressive individuals. In addition, it can be argued that it might be difficult to find effects on outcome measures of aggression in a relatively non aggressive sample such as we used here. Somewhat related to this, the current study didn't include measures of pre-existing hostile schemas of the participants. Considering previous findings that suggest that maladaptive attention allocation may only be related to aggression in individuals who hold hostile schema (e.g., Troop-Gordon et al., 2018), it is possible that the current training is only beneficial for individuals holding negative perceptions of others. Second, the measure of AB did not correlate significantly with the concurrently assessed aggression related measures, raising some questions about the validity of the currently adopted approach to assessing aggression related attention bias. Interestingly, a recent study did find a significant relation between a measure of aggression and a gaze pattern that somewhat similar to the one we used to operationalize AB in this study. That is, Laue et al. (2018) showed participants 3 image cartoon stories in which the first image illustrated the context, the second picture showed one character doing something that negatively affected another character, and a third picture

showing the negative outcome and the facial expression of the harm-doer. The sequence of presentation was such that image 1 and 2 were subsequently presented alone, and then image 2 and 3 were presented on screen together. Results showed that when the final two pictures were presented together, individuals with higher aggression scores tended to look longer at the negative act in picture 2 than at the facial expression of the harm-doer in picture 3. Thus, higher aggression seemed to be related to more attention for the negative event than potentially mitigating information from the facial expression of the harm-doer. This is rather similar to our operationalization of attentional bias: more attention to the negative outcome of the incident than the facial expression of the harm-doer. However, one difference is that Laue et al. 2018 studied attention to the negative *act*, while in the current study we focused on the *outcome* of the act. Future work could explore whether this difference can explain why Laue and colleagues (2018) did and we did not find a relationship with a measure of aggression. Third, the lack of a control group means that we cannot completely preclude the possibility that the positive change in attention bias is due to some other factors. Future research needs to compare the positive training to a control group with a neutral training in order to more rigorously test its effectiveness on attentional processes. Fourth, although we have demonstrated the possibility that a single session of positive attentional training using gaze-contingencies could induce attention bias to pro-social cues, the training did not differentially affect aggression-related measures. Therefore, a possible related limitation might have to do with the number of sessions and trials of the training. In our study, participants completed a total number of 40 training trials during a single-session. Previous gaze-contingent studies showed a large variation in number of trials and sessions (e.g., Ferrari et al., 2016; Price et al., 2016; Sanchez, Everaert & Koster, 2016; Lazarov et al., 2017). Despite those variations between studies, the results showed that the training was successful in changing gaze patterns in the intended direction. However those training effects differed in regards to symptom reductions. Single-session studies with a high number of trials (i.e., 270 trial), have found no changes in mood in response to a stressor (e.g., Ferrari et al., 2016). Single-session studies with a lower number of trials (i.e., 48 trials), were found to be successful in reducing negative emotions (e.g., Sanchez et al., 2016). On the other hand, previous gaze-contingent studies using even less trials (i.e., 30 trials) with a higher number of sessions (i.e., 8 sessions) found a great symptom reduction in socially anxious participants (e.g., Lazarov et al., 2017). This might suggest that future

gaze-contingent attention training methodologies with limited number of trials might benefit from increasing the number of training sessions to produce higher impact on symptom reduction.

Additionally, because participants were explicitly instructed to attend to the information that indicated whether the incident happened on purpose or not, and because they received feedback on their response (the cue they payed attention to) during the training, we may not only have trained attention deployment, but also participants' interpretation of the cues in the social situations. At this point it is impossible to disentangle these possible effects. It is interesting to note, however, that participants' interpretations of the situations became significantly more pro-social after training in *both* training conditions. While this indicates that, as discussed above, the direction of attention did not have the expected effect in this study, the observed effects might be due to our instructions that were aimed at improving encoding of social cues. Perhaps making participants more aware of what they are looking at to decide whether something happened on purpose or not was sufficient to alter interpretations, regardless of the direction of attention. At this time, this is speculation, however, future research should include a neutral training condition to ensure that the observed changes were due to the training and not simply test-retest effects. Finally, in order to further the potential effectivity of the present CBM-A gaze-contingent training in modifying attention bias over other existing attention training methodologies such as dot-probe task, future research should directly compare the two methodologies.

To conclude, this is one of the first studies that developed and tested a novel gaze-contingent procedure targeting attention in the context of aggression bias. Importantly, our study shows that a single session of this novel gaze-contingent CBM-A was able to modify attention bias in a pro-social direction. However, we did not find evidence for effects of the training on interpretation bias, aggressive behavior and mood. That being said, the training is still in its early stages and as discussed above there are a number of aspects of the training that might be adjusted in order to get the desired effects on aggression. We hope that future research will further explore and improve the potential impact of this training on the attentional processes underlying aggression, and aggressive behavior.

# CHAPTER 6

## General Discussion

Advances in understanding the role of cognitive biases of attention and interpretation in aggression have led to a significant new interest in applying this knowledge to intervention research (de Castro, Veerman, Koops, Joop, & Monshouwer, 2002; Troop-Gordon, Gordon, Vogel-Ciernia, Lee, & Visconti, 2018). Results of the first intervention studies have shown that modifying a specific information processing bias using the cognitive bias modification (CBM) paradigm could result in the modification of the targeted bias and significant anger and aggression reduction (e.g., Hawkins & Cougle, 2013; Vassilopoulos, Brouzos & Andreou, 2015). However, at the start of this dissertation project, aggression studies were limited in their examination of the effects of CBM paradigms targeting interpretations (CBM-I). Additionally, there had not been any studies on the effectiveness of cognitive bias modification paradigms targeting attention (CBM-A) on both bias and symptom reduction in the context of aggression.

Moreover, in aggression studies, interpretation biases were typically assessed and trained in isolation using vignettes describing hypothetical provocative social situations in which the intention of the harm-doer is ambiguous (e.g., Hawkins & Cougle, 2013; Vassilopoulos et al., 2015). However, in real-life situations, visual nonverbal cues such as facial and physical expressions carry important signs regarding the internal state of others, including their intentions (Cadesky, Mota, & Schachar, 2000). This raises the question of whether including visual ambiguous social scenes in the training procedure might provide a more information-rich and naturalistic context, which increases the effect of the training procedure on both interpretation and attention biases. Therefore, the aim of the present dissertation was to examine whether a novel CBM procedure using pictorial stimuli can be used to modify attention and interpretation biases in the context of aggression and examine the effect of those modifications on aggression.

Additionally, the cognitive bias hypothesis suggests that cognitive biases are interrelated and that training procedures that target a combination of biases have a greater impact on disorders than targeting a cognitive process in isolation (Hirsch, Clark, & Mathews, 2006). This raises the question of how cognitive biases of attention and interpretation might interact and contribute to aggressive behavior. Also, it raises the question whether there is an added value of a combined bias training targeting both attention and interpretation bias relative to a single bias training in both bias and aggression reduction.

Thus, in order to examine whether it is possible to modify aggression by modifying one's attention and interpretation biases and examine the use of pictorial stimuli in boosting more training effects, the present research project focused on answering the following four questions: (1) Can a novel CBM training procedure using pictorial stimuli be used to change interpretation and attention biases in the context of aggression?; (2) Do changes in attention or interpretation biases lead to changes in aggression?; (3) How do attention and interpretation biases interact in maintaining aggression?; and (4) Is a combined bias CBM training procedure more effective than a single bias CBM training procedure on both bias and aggression reduction?



In order to investigate those questions, we first examined whether a novel CBM of interpretations (CBM-I) paradigm using pictorial stimuli can be used to modify interpretation bias, and we examined the effect of those changes on aggression (Chapter 2). Second, we tested whether a novel gaze-contingent CBM of attention (CBM-A) using pictorial stimuli can be used to modify attention bias, and we then examined the effect of those changes on aggression. Furthermore, we examined the interrelation between both attention and interpretation bias by examining the effect of CBM-A on interpretation bias (Chapter 3). Third, we investigated whether we could replicate our findings regarding the effect of CBM-I on interpretation bias and aggression, and we additionally explored the effect of CBM-I on attention bias and interpretation bias of facial expressions (Chapter 4). Lastly, we examined the effectiveness of a combined CBM of attention and interpretation bias (CBM-AI) on both bias and aggression reduction and compared those results to a single CBM targeting interpretations and a control condition (Chapter 5).

The main hypothesis underlying these four studies is that CBM procedures in the context of aggression can modify the targeted bias in the intended direction and thereby affect aggression. Additionally, based on the cognitive bias hypothesis (Hirsch et al., 2006), we expected the modification of one cognitive bias to influence other untrained cognitive biases. Lastly, we expected that a CBM targeting both attention and interpretation bias would produce greater aggression reduction than training a single bias in isolation.

In this final chapter, the findings of the four research questions are discussed and reviewed. Furthermore, I address theoretical and methodological implications, limitations, and directions for future research.

**Can a novel CBM training procedure using pictorial stimuli be used to change interpretation and attention biases in the context of aggression?**

As a first step to address this question, the pilot study described in Chapter 2 examined whether a novel CBM-I procedure using pictorial stimuli influences interpretations. We expected that training individuals to interpret ambiguous situations in a pro-social way would increase pro-social interpretation bias, whereas training them to interpret such situations as hostile would increase hostile interpretation bias. The results showed that a single session of positive interpretation training using pictorial stimuli indeed increased pro-social interpretation bias. This finding is well in line with previous studies that suggested that interpretation biases can be trained in a pro-social way (Hawkins & Cougle, 2013; Vassilopoulos et al., 2015). However, the negative training did not result in an increase in hostile interpretation bias. This is in contrast to previous findings that have shown that it is possible to induce hostile interpretation bias (e.g., Hawkins & Cougle, 2013). However, in our study participants in the negative training were less accurate in following the training instructions than participants in the positive training condition. Participants who followed the instructions and performed better in the negative training condition (i.e., lower error rate) showed significant change in their interpretation bias in a hostile direction.

One of the main limitations of this study was the lack of a control group, which means that we were not able to conclude that the positive change in interpretation bias was completely due to our CBM-I training procedure. To address this issue, the studies described in Chapter 4 and 5 included a control condition which provided participants with a task identical to the task of participants in the other training conditions, but was not intended to cause any changes in either attention or interpretation bias of the social cues. In both studies we trained participants using a modified version of our CBM-I, and the results of the study described in Chapter 4 showed that the training changed measures of interpretation bias in a pro-social direction compared to a control training. In addition, the negative CBM-I training changed measures of interpretation bias in a hostile direction but not more so than a control training. The findings described in Chapter 5 showed that participants' pro-social interpretation bias increased from pre- to post-training. However, due to methodological issues related to the control condition (described in Chapter 5), the changes in pro-social interpretation bias were not different than in the control condition. Nevertheless, the effects of our CBM-I on pro-social interpretation bias were consistent across all

three studies, which seems to warrant the conclusion that our training procedure using pictorial stimuli can successfully induce more pro-social interpretation processing styles in the context of aggression.

Next, the study described in Chapter 3 examined whether our novel CBM procedure using pictorial stimuli influences attention bias. Findings from eye-tracking studies suggested that aggression is associated with a specific pattern of inattention to relevant social cues that is guided by pre-existing hostile schemas (Horsley, de Castro, & van der Schoot, 2010; Troop-Gordon et al., 2018). Schemas are internal representations of previous social interactions that are stored in the memory and may influence the processing of future social cues (de Castro & van Dijk, 2018). For example, one study showed that aggressive children who hold negative peer beliefs evinced initial inattention to social cues and poor recall of those cues (Troop-Gordon et al., 2018). This indicates that pre-existing hostile schemas in combination with an inattention to relevant social cues may lead to overreliance on those hostile schemas when interpreting ambiguous social situations. Following this, aggressive individuals might benefit most from training programs that train them to effectively attend to and encode relevant social cues that would help disambiguate the social situation. We argued earlier that the individual's facial expressions hold informative cues regarding their intentions. Thus, we wanted to know whether training participants to attend more to the face of the harm-doer (i.e., adaptive cues) and *not* to the negative outcome would help disambiguate the situation in a pro-social manner even when the facial expression was ambiguous.



In order to do so, we investigated the efficacy of a novel gaze-contingent CBM-A procedure using pictorial stimuli, aimed at training more adaptive (i.e., helpful in disambiguating the situation) looking strategies in the context of aggression. Results indicated that gaze-contingent CBM-A training within the positive condition indeed resulted in an increase in attention to adaptive cues (i.e., the face of the harm-doer) than to maladaptive cues (i.e., the negative outcome of the situation) in a picture of an ambiguous social situation. These findings are in line with previous studies that suggested the potential efficacy of using a gaze-contingent approach in modifying attentional processes by inducing a positive bias (Ferrari, Mobius, van Opdorp, Becker, & Rinck, 2016; Price, Greven, Siegle, Koster, & De Raedt, 2016; Lazarov, Pine, & Bar-Haim, 2017; Sanchez, Everaert, & Koster, 2016). We assume that this increase in adaptive attention bias found in our training is related to methodological advantages of the gaze-contingent CBM-A

paradigm. In our CBM-A training, participants had to fixate and maintain their attention on a specific cue on a visual stimuli. This setup allowed for a more direct training of gaze direction using feedback. Also, contrary to the probe detection paradigms, this setup allows the use of a more complex and representative selection of visual stimuli that a normal person would encounter in their day-to-day life. Surprisingly, however, our CBM-A training was not successful in increasing attention bias to maladaptive cues (i.e., the negative outcome), as no changes in attention bias were found in the negative condition. Regarding this result, it is important to note that before training, our sample did not show the attention bias toward adaptive cues which we expected for healthy participants. Thus, a possible floor effect may explain the lack of effect of CBM-A training in the negative condition. One explanation for this pre-existing "maladaptive" attention bias may be that negative cues are more salient and attention-grabbing than pro-social cues in social situations (Wadlinger & Isaacowitz, 2008). Findings by Ferrari et al. (2016) showed that before the gaze-contingent CBM-A training, participants took longer to disengage from negative than positive stimuli, which provides further support to our explanation of why we were not able to further increase attention bias to maladaptive cues. To our knowledge, this was the first study to train attention bias using gaze-contingencies in aggression, and the results indicate the efficacy of the gaze-contingent CBM-A procedure to modify attention bias in an adaptive way.

Overall, the current findings suggest that our novel CBM training procedures can indeed change attention (Chapter 3 and 5) and interpretation (Chapter 2, 4 and 5) biases in a pro-social way. After establishing the effectiveness of our CBM-I and CBM-A training in modifying the targeted bias, the crucial next step would be to examine the effects of those modified cognitive biases on aggression.

**Do changes in attention or interpretation biases lead to changes in aggression?**

In addition to examining the effect of CBM-I and CBM-A training on bias reduction, we measured whether these CBM training paradigms would affect aggression. Our main hypothesis was that CBM-I and CBM-A would lead to changes on the targeted bias, which in turn would lead to changes in aggression depending on the training condition. In the studies that included the CBM-I training (Chapter 2, 4, and 5), we expected that an increase in pro-social interpretation bias would lead to a reduction in aggression and that an increase in hostile interpretation bias would lead to an increase in aggression.

To begin with, the results of the pilot study described in Chapter 2 were promising. The CBM-I training did not only result in an increase in pro-social interpretation bias but also led to a reduction in verbal aggression and self-reported anger and an increase in self-reported happiness. Moreover, in the negative training, the results showed that the more participants' interpretations changed in a hostile way, the more they responded aggressively on the Taylor Aggression Paradigm (TAP; Taylor, 1967). These results are consistent with past studies demonstrating that modifying interpretation bias may affect aggression (e.g., Hawkins & Cougle, 2013; Vassilopoulos et al., 2015) and change mood (Lothmann, Holmes, Chan, & Lau, 2011). Unfortunately, we were not able to replicate those results in the studies described in Chapter 4 and 5, in which the change in interpretation bias did not affect aggression nor mood in the expected direction.



The study described in Chapter 3 examined the effect of a gaze-contingent CBM-A training on aggression. We expected that training participants who attended to maladaptive cues would show an increase in subsequent maladaptive attention bias and thereby an increase in aggressive behavior. Conversely, we expected that training them to attend to adaptive cues would increase adaptive attention bias and reduce subsequent aggressive behavior. However, both training conditions did not result in changes in aggression and mood in the expected direction. These results are in contrast to studies in the context of anxiety, which suggested that gaze-contingent CBM-A training leads to symptom change (e.g., Lazarov et al., 2017).

Overall, the effects of the CBM-I on aggression in the current dissertation were inconsistent, and we found no evidence for the efficacy of CBM-A on aggression. For this reason, we are not able to draw any firm conclusions regarding the efficacy of our CBM training paradigm on aggression. An explanation for these unexpected findings is that in all our studies we included healthy university students, which were low in state and/or trait aggression, making it more difficult to find reductions in our aggression measures due to floor effects. Another explanation is that the behavioral aggression measures that were used in the current dissertation to assess aggression post-CBM training were not always optimally suited to assess the effect of the training procedures and not sensitive to the type of change in aggression caused by the training. In both studies described in Chapter 2 and 3, we used a competitive reaction time game where the participant believes that he or she is playing against an opponent (i.e., TAP; Taylor, 1967). The measure assess the intensity and duration that the participant aggresses against their opponent

during the game. During this task the intention of the opponent can be interpreted in a pro-social and hostile way. Therefore task performance may be related to changes in interpretations over training as were described in Chapter 2. However, responses in the TAP are probably independent of, and thus not influenced by attending to adaptive or maladaptive cues, because the task does not include such cues. This might explain why we found effect of the CBM-I (in Chapter 2) but not the CBM-A (in Chapter 3) on the TAP.

In the study described in Chapter 5, we used a slightly different way of measuring aggression by creating a provocative situation using a joystick game task (i.e., Technical Provocation Paradigm (TPP); Panagiotidis et al., 2017). In this task, aggression was measured by how hard participants pulled the joystick toward themselves. Importantly in this task aggression is not called up in a social interaction. Thus, performance on the TPP task is not dependent on attention to specific social cues or the interpretation of an ambiguous social situation, and may therefore not be affected by the training. Thus, the most straightforward explanation for the lack of the CBM training effect on these behavioral measures (in studies described in Chapter 3 and 5) is a mismatch between the type of processes we trained and the type of processes underlying responses on the specific aggression task. In line with this suggestion, Hawkins and Cougle (2013) found that induced pro-social interpretation bias was successful in reducing anger reactivity to an interpersonal insult, a situation in which interpretations are probably crucially important for the individual's response. Thus, it is possible that the CBM has a genuine effect specifically on interpersonal aggression-related context, and that the most efficient way to measure the effects of CBM training on aggression is by using a task that also implies the targeted bias.

Nevertheless, research in this area is still scarce and more research is still needed to determine the efficacy of CBM-A and CBM-I training paradigms on aggression reduction.

**How do attention and interpretation biases interact in maintaining aggression?**

The Social Information Processing (SIP) model suggests that biases in attention and interpretation are interrelated rather than independent (Crick & Dodge, 1994). This raises the question of how these biases may interact and contribute to aggressive behavior. Interestingly, previous anxiety studies have demonstrated that modifying one cognitive bias has a significant effect on another cognitive bias, suggesting that these cognitive processes are not independent but influence one another in maintaining anxiety (e.g., Amir, Bomyea, & Beard, 2010; White, Suway,

Pine, Bar-Haim, & Fox, 2011). Knowing whether modifying one cognitive process (e.g., attention) is sufficient to alter another (e.g., interpretation) provides interesting implications for aggression interventions that target multiple biases. CBM training paradigms allow us to better understand how these biases interact in the context of aggression, by modifying one bias and assessing the impact on the other. However, to date, no aggression studies have focused on the question of whether modification of interpretation bias can influence attention bias and vice versa.

As a first step to address this gap in the literature, the study described in Chapter 3 focused on the question of whether inducing changes in attention bias using CBM-A would influence how subsequent ambiguous social information is interpreted. We expected that participants who were trained to attend to adaptive cues would make less hostile interpretations than participants who were trained to attend to maladaptive cues. Contrary to previous anxiety research (Bowler et al., 2017; White et al., 2011), the attentional training did not lead to changes in interpretation bias in the expected direction. That is, in both training groups, interpretation bias increased in a pro-social way and thus seemed independent of the induced attention bias. Possibly other factors, such as specific procedural details of the training paradigm (e.g., instructions, training setting, and/or stimulus material), may have caused the effects on interpretation bias.

Next, in the studies described in Chapter 4, we examined the effect in the opposite direction and tested the impact of training interpretation bias using CBM-I on attention bias. We expected that increasing pro-social interpretation bias would lead to heightened attention to adaptive cues, and that increasing hostile interpretation bias would lead to heightened attention to maladaptive cues. Results showed that in all training conditions, attention bias changed in a pro-social direction. One explanation may be that the modest changes in interpretation bias were not strong enough to lead to changes in attention bias in the expected direction. In the study described in Chapter 5, we again addressed the relation between interpretation and attention bias using a CBM-I training procedure in one of the training conditions. The results indicated that CBM-I led to an increase in both pro-social interpretation and attention bias from pre- to post-training. Contrary to our expectations, those changes were not different than in the control condition. However, a number of methodological issues related to the control condition (described in Chapter 5) are probably the cause of changes in interpretation bias.

How might changes in interpretation bias (in the study described in Chapter 5) lead to changes in attention bias but not vice versa (the study described in Chapter 3)? There are a number of possible explanations for this. First, the current findings might fit with the reasoning of Horsley et al. (2010) and Troop-Gordon et al. (2018) that aggressive individuals rely on pre-existing hostile schemata when interpreting a social situation rather than attending to relevant social cues in the situation. Thus, the increase in pro-social interpretation bias (in the study described in Chapter 5) may have helped participants in accessing more pro-social expectations of the situation. This may have made it easier for participants to attend to social cues (i.e., adaptive cues) that fit their pro-social expectations of the situation.

Second, the available evidence seems to suggest that interpretation training might have a stronger effect influencing other forms of biases (e.g., attention and interpretation of facial expression) then attention training. Yet, it might be that the effects of some training-specific procedural details (e.g., training instructions, presentation time, or feedback) may elicit the efficacy of the CBM training paradigms, which might specifically explain the increase in both attention and interpretation bias in the control condition (in Chapter 5). If this indeed would be the case, the crucial next step would be to test the key procedural details influencing the CBM training outcomes in the context of aggression. Also, future work should include a truly neutral control condition before drawing any firm conclusions related to the interrelation between attention and interpretation biases.

In the study described in Chapter 4, our aim was to extend the findings of Chapter 2 by answering the question of whether the CBM-I training also impacted participants' interpretation of facial expressions. We mentioned earlier that visual nonverbal cues such as facial and physical expressions carry important signs regarding the internal state of others including their intentions (Cadesky et al., 2000). Thus, we wanted to know what other factors may contribute to the training effects, which might provide cues as to how the effects of the CBM-I training might be strengthened. In order to do so, in the study described in Chapter 4, we examined the effect of the modified interpretation bias of intent and whether it influenced the interpretation bias of facial expression. We expected that training individuals to interpret ambiguous situations as pro-social would lead to an increase in pro-social intent attribution, which in turn would increase pro-social interpretation of facial expressions, whereas training them to interpret such situations as hostile

would increase hostile intent attribution bias, which in turn would increase hostile interpretation of facial expressions. The results of the study indicated that due to the modest effect of the CBM-I on the interpretation bias of intent, the transfer effect of the modified interpretation of intent to the interpretation of facial expression may have been limited. Therefore, we wanted to further examine the relation between interpretation of intent and interpretation of facial expression.

Thus, in the study described in Chapter 5, we again measured the interpretation of facial expressions. Interestingly enough, the results indicated that in all training conditions not only attention and interpretation bias changed in a pro-social direction but pro-social interpretation of facial expressions also changed in a pro-social direction from pre- to post-training. Although we are not aware of previous studies that examined the relation between modified cognitive bias of interpretation and its relation to interpretation of facial expressions, there is one study that examined the effect of modified hostile interpretations of facial expressions on interpretations of intent in an aggressive sample using morphed faces. The results of that study showed that the reduction in hostile interpretation of facial expressions did not generalize to changes in participants' interpretation of intent (Hiemstra, de Castro, & Thomaes, 2018). The authors argued that CBM is more effective in modifying the targeted bias than other forms of biases. Interestingly, our findings suggest that modification of interpretation bias of intent may have an effect on influencing other (related) forms of biases. However, at this stage this is only speculation, and the association between biases of intent attribution and interpretation of facial expression should still be tested in future work.

Overall, the current findings suggest that biases in interpretation and attention can interact in the context of aggression. Specifically, the current dissertation provides initial results indicating that the way that individuals interpret the intentions of others affects their attention allocation to social cues and how they interpret ambiguous facial expressions. Thus, knowing that training interpretation bias of intent has the same effect on biases of attention and interpretation of facial expressions, might have useful practical implications for aggression reduction interventions targeting multiple cognitive biases.

**Is a combined bias CBM training procedure more effective than a single bias CBM training procedure on both bias and aggression reduction?**

Even though we could not find clear evidence of the interrelation between attention and interpretation bias, it might still be the case that training both these biases maximizes the effect on aggression reduction. The study described in Chapter 5 explored the added value of a combined CBM targeting both attention and interpretation bias (CBM-AI) relative to a single bias training (e.g., CBM-I). The results of the study indicated that a single-session of CBM-AI indeed increased attention allocation to adaptive cues and pro-social interpretation bias of intent. However, contrary to our expectations, we found that those changes were not significantly different from changes in the CBM-I and control conditions. The only previous study (i.e., social anxiety) we are aware of that compared the combined bias training to single bias training and control condition also did not find combined bias training to be more effective in bias change than the other training conditions (Naim, Kivity, Bar-Haim, & Huppert, 2018). Although the participants in the control condition were not meant to be trained, it is possible that the observed changes in both attention and interpretation bias were due to the fact that their interpretations were trained (i.e., due to methodological issues reported in Chapter 5). Moreover, given the fact that attention bias was only trained in the CBM-AI training condition, we suggested that changes in attention bias across all training conditions were due to the increase in pro-social interpretation bias from pre- to post-training and that those changes generalized to attention bias. This suggestion is well in line with previous anxiety studies that support the impact of induced positive interpretation bias on attentional avoidance to threat stimuli (e.g., Amir et al., 2010; Mobini et al., 2014). However, we mentioned previously that in our CBM-A training (Chapter 3), the increase in pro-social attention bias did not impact interpretation bias. Additionally, our CBM-I training (Chapter 4) resulted in significant differences between the groups at post-training, while the changes in interpretation bias from pre- to post-training were not significant. Additionally, the study showed no changes in measures of interpretation bias of facial expression and perceived anger. However, since interpretation bias changed significantly from pre- to post-training in the study described in Chapter 5, we found a transfer of effect not only to attention bias but also interpretation bias of facial expression and perceived anger, which changed in a positive way. This might suggest that in the context of aggression, interpretation bias may have more of a generalizable cognitive effect on other biases compared to attention bias. Additionally, in contrast to the findings of the study

described in Chapter 4, the stronger changes in interpretation bias may have had a stronger effect on attention, interpretation bias of facial expressions, and perceived anger. Nonetheless, future work is clearly needed to confirm these findings.

The results described in Chapter 5 also showed that none of the training conditions showed an effect on aggression measures. These results are in contrast to previous anxiety findings, which found a reduction in state and trait anxiety following a combined bias training targeting both attention and interpretation bias (e.g., Brosan, Hoppitt, Shelfer, Sillence, & Mackintosh, 2011; Beard, Weisberg, & Amir, 2011; Lisk, Pile, Haller, Kumari, & Lau, 2018). The only study that compared a combined bias training to a single bias training and a control condition did not find the combined bias training to be more effective than the single bias training (Naim et al., 2018).



Taken together, we have no evidence to draw firm conclusions about whether a combined bias CBM training procedure is more effective than a single bias CBM training procedure in both bias and aggression reduction.

**Theoretical and methodological implications**

This dissertation contributes to the literature of CBM paradigms in the management of aggression. First, the study described in Chapter 2 provides evidence for the efficacy of a novel training paradigm in modifying interpretation bias in a pro-social way using pictorial stimuli. Additionally, our findings add to previous aggression studies demonstrating that interpretation bias can be modified (e.g., Hawkins & Cougle, 2013; Vassilopoulos et al., 2015). However, most experimental CBM studies in aggression focus on CBM-I, while studies into the efficacy of CBM-A lag behind. Thus, the study presented in Chapter 3 is the first to show that it is possible to modify attention bias in an adaptive way using a gaze-contingent CBM-A procedure with the use of pictorial stimuli. Additionally, these findings add to the other gaze-contingent CBM-A studies that provided evidence of the feasibility of using a gaze-contingent approach to train attentional deployment (e.g., Ferrari et al., 2016; Price et al., 2016; Sanchez et al., 2016; Lazarov et al., 2017).

Second, the studies in this dissertation contribute to the combined cognitive bias hypothesis (Hirsch et al., 2006) by first, providing knowledge regarding the interrelation between attention and interpretation bias (studies presented in Chapter 3, 4 and 5). The results of CBM training presented in Chapter 3, 4 and 5 suggest that only when interpretation bias changes significantly

from pre- to post-training it would lead to changes in attention bias but not vice versa. Related to this, our study further expands the aggression-related interpretation bias by investigating the relations between interpretation of intent and interpretation of facial expressions. Our findings suggest that the interpretation of intent and the interpretation of facial expressions are closely related. Additionally, modifying interpretation bias of intent using a CBM-I paradigm may influence interpretation bias of the facial expressions of others. These results are in contrast with the findings of Hiemstra et al. (2018), who found that the modification of interpretation bias of facial expressions did not generalize to changes on participants' interpretation of intent. Thus, our results seem to indicate that changes in interpretation bias can generalize to affect other forms of biases but not vice versa. However, this possible effect of interpretation bias in the context of aggression would still need further study to confirm these findings.

Second, our findings contribute to the combine bias hypothesis by examining the effect of a combined bias training relative to a single bias training (study presented in Chapter 5). Our findings illustrate that the combined bias CBM training targeting attention and interpretation bias is not more effective than a single bias training targeting interpretations. This finding is consistent with the findings of Naim et al. (2018), who found that the combined bias training was not more effective than a single bias training in the context of anxiety. However, due to the problems with the control condition described in Chapter 5, the results regarding the efficacy of the combined bias training are inconclusive, and more research is still needed.

**Limitations and direction for future research**

Given the novelty of our CBM training procedure, we included a sample of healthy university students across all studies described in this dissertation, in order not to prematurely burden a clinical sample with untried procedures. While this was a deliberate choice, including a healthy sample makes it more difficult to discover the effects on outcome measures of aggression. However, previous studies have found significant relations between measures of aggression and cognitive biases in similar samples in the past (see Klein Tuente, Bogaerts, & Veling, 2019 for a recent review) and we expected that the modified cognitive biases would lead to changes in aggression among healthy participants. Nevertheless, in most of our studies the CBM training did not lead to the expected effect on aggression measures. As a result, it precludes any firm conclusions regarding the efficacy of our CBM on aggression.

Another issue, which may also be related to the lack of effect of our training procedures on aggression, is that we only used single session training procedures with a limited number of trials. This may have resulted in insufficiently large changes in the targeted biases to affect aggression. That is, across all studies in the current dissertation, participants completed a total number of 40 training trials during a single session. Previous experimental studies of CBM studies showed a large variation in number of trials and sessions, and there is no indication of the number of trials and sessions necessary to achieve both bias and symptom change. However, it has been suggested that multiple-session training might be necessary for greater change and long-term durability on both cognitive bias and symptom reduction (Hallion & Ruscio, 2011). Since 40 training trials led to bias but not aggression change, this might suggest that future CBM training methodologies with limited numbers of trials might benefit from increasing the number of training sessions to produce a higher impact on both bias and aggression reduction.



A third issue is that we did not take into account the potential influence of offline processes on our CBM training. For instance, the SIP model emphasizes the role of schemas on the individual's social processing, which may influence which cues they attend to and how those cues are interpreted (Crick & Dodge, 1994). Normally, social situations consist of varying social cues, and the role of social schemas is to facilitate the processing of those cues (Simons & Burt, 2011). Additionally, similar to schemas, normative beliefs may affect the individual's social information processing and might play a role in activating the appropriate schema (Crick & Dodge, 1994). Normative beliefs refer to the individual's own judgment about which behaviors are considered acceptable and which are not (Huesmann & Guerra, 1997). For instance, the belief that aggression is an acceptable response in social interactions might increase the chance of activating schemas emphasizing aggressive responses, which may in turn increase the likelihood of interpreting someone else's behavior as hostile and responding aggressively (Werner & Nixon, 2005). Thus, future studies should examine the role of these offline processes in the efficacy of CBM-A or CBM-I on both bias and aggression reduction. It is possible that our CBM training is more beneficial for individuals with a pre-existing hostile schema and/or normative beliefs about aggression.

A fourth issue is that it is unclear whether the modified biases of attention and interpretation in the current dissertation would have generalized to similar, but distinct, measures. In the study

described in Chapter 4, we included a different measure for interpretation bias. However, the results of the CBM-I training did not transfer to this different interpretation bias measure. More research is needed to determine whether and, if so, how future CBM training paradigms can be improved to have more generalizable effects across different tasks.

Finally, in the study described in Chapter 2, we suggested that using visual stimulus materials to modify interpretation bias and to use a gaze-contingent procedure to modify attention bias might be more powerful in bias manipulation and boost the training effect. Since we did not compare our training CBM methods to other existing methods, we could not conclude the potential superiority of our training paradigms. Thus, future studies could further validate our training and test its additive efficacy over other existing CBM-A tasks, such as the dot-probe or CBM-I tasks, such as written vignettes, and directly compare between them.

**Conclusion**

Based on the SIP model (Crick & Dodge, 1994), the studies in this dissertation investigated whether both novel CBM procedures (CBM-I using pictorial stimuli designed to modify interpretation bias and gaze-contingent CBM-A using pictorial stimuli designed to modify attention bias) would result in a change in the targeted cognitive bias and aggression. Based on the results described in this dissertation, several conclusions can be drawn. First, our findings show that each CBM was successful in modifying its targeted bias in a pro-social way. Second, our CBM training procedures were not effective in reducing aggression or negative mood. Third, we were not able to find evidence of the interrelation between attention and interpretation bias in the context of aggression. Some of our findings, however, may suggest that compared to attention bias, interpretation bias has a more generalizable effect in the context of aggression. Finally, we found that a combined bias training does not have an added efficacy over a single bias training in aggression reduction. Overall, the current dissertation adds knowledge related to the role of attention and interpretation bias in the context of aggression, but a number of issues remain inconclusive, and more research is needed. We hope that the studies described in this dissertation will inspire further research on the CBM training paradigms that can help find the most effective way of producing positive changes in cognitive information processing underlying aggression.

# R

## References

AlMoghrabi, N., Huijding, J., & Franken, I. H. (2018). The effects of a novel hostile interpretation bias modification paradigm on hostile interpretations, mood, and aggressive behavior. *Journal of Behavior Therapy and Experimental Psychiatry*, *58*, 36-42.

AlMoghrabi, N., Huijding, J., Mayer, B., & Franken, I. H. (2019). Gaze-contingent attention bias modification training and its effect on attention, interpretations, mood, and aggressive behavior. *Cognitive Therapy and Research, 43*, 861-873.

Amir, N., Beard, C., Burns, M., & Bomyea, J. (2009). Attention modification program in individuals with generalized anxiety disorder. *Journal of Abnormal Psychology, 118*, 28-33.

Amir, N., Beard, C., Taylor, C. T., Klumpp, H., Elias, J., Burns, M., & Chen, X. (2009). Attention training in individuals with generalized social phobia: A randomized controlled trial. *Journal of Consulting and Clinical Psychology*, *77*, 961-973.

Amir, N., Bomyea, J., & Beard, C. (2010). The effect of single-session interpretation modification on attention bias in socially anxious individuals. *Journal of Anxiety Disorders*, *24*, 178-182.

Anderson, C. A., & Bushman, B. J. (2002). Human aggression. *Annual Review of Psychology*, *53,* 27-51.

Bar-Haim, Y. (2010). Research review: Attention bias modification (ABM): A novel treatment for anxiety disorders. *Journal of Child Psychology and Psychiatry*, *51*, 859-870.

Beard, C., Sawyer, A. T., & Hofmann, S. G. (2012). Efficacy of attention bias modification using threat and appetitive stimuli: A meta-analytic review. *Behavior Therapy*, *43*, 724-740.

Beard, C., Weisberg, R. B., & Amir, N. (2011). Combined cognitive bias modification treatment for social anxiety disorder: A pilot trial. *Depression and Anxiety*, *28*, 981-988.

Boffo, M., Zerhouni, O., Gronau, Q. F., van Beek, R. J., Nikolaou, K., Marsman, M., & Wiers, R. W. (2019). Cognitive bias modification for behavior change in alcohol and smoking addiction: Bayesian meta-analysis of individual participant data. *Neuropsychology Review*, *29*, 52-78.

Bowler, J. O., Hoppitt, L., Illingworth, J., Dalgleish, T., Ononaiye, M., Perez-Olivas, G., & Mackintosh, B. (2017). Asymmetrical transfer effects of cognitive bias modification: Modifying attention to threat influences interpretation of emotional ambiguity, but not vice versa. *Journal of Behavior Therapy and Experimental Psychiatry*, *54*, 239-246.

Brosan, L., Hoppitt, L., Shelfer, L., Sillence, A., & Mackintosh, B. (2011). Cognitive bias modification for attention and interpretation reduces trait and state anxiety in anxious patients referred to an out-patient service: Results from a pilot study. *Journal of Behavior Therapy and Experimental Psychiatry*, *42*, 258-264.

Brown, S. C., & Craik, F. I. M. (2000). Encoding and retrieval of information. In E. Tulving, & F. I. M. Craig (Eds.), *The Oxford handbook of memory*, (pp. 93-107). Oxford, UK: Oxford University Press.

Brugman, S., Lobbestael, J., Arntz, A., Cima, M., Schuhmann, T., Dambacher, F., & Sack, A. T. (2015). Identifying cognitive predictors of reactive and proactive aggression. *Aggressive Behavior*, *41*, 51-64.

Buss, A. H., & Perry, M. (1992). The aggression questionnaire. *Journal of Personality and Social Psychology, 63*, 452-459.

Cadesky, E. B., Mota, V. L., & Schachar, R. J. (2000). Beyond words: how do children with ADHD and/or conduct problems process nonverbal information about affect?. *Journal of the American Academy of Child & Adolescent Psychiatry*, *39*, 1160-1167.

Cohen, D. J., Eckhardt, C. I., & Schagat, K. D. (1998). Attention allocation and habituation to anger-related stimuli during a visual search task. *Aggressive Behavior*, *24*, 399-409.

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.

Curtis, D. S., Epstein, N. B., & Wheeler, B. (2017). Relationship satisfaction mediates the link between partner aggression and relationship dissolution: The importance of considering severity. *Journal of Interpersonal Violence*, *32*, 1187-1208.

de Castro, B. O. (2004). The development of social information processing and aggressive behaviour: Current issues. *European Journal of Developmental Psychology*, *1*, 87-102.

de Castro, B. O., & van Dijk, A. (2017). "It's gonna end up with a fight anyway": Social cognitive processes in children with disruptive behavior disorders. In J.E. Lochman & W. Matthys (Eds.), *The Wiley handbook of disruptive and impulse-control disorders*, (pp. 237-253). Hoboken, NJ: John Wiley & Sons Limited.

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.

Dillon, K. H., Allan, N. P., Cougle, J. R., & Fincham, F. D. (2016). Measuring hostile interpretation bias: the WSAP-hostility scale. *Assessment*, *23*, 707-719.

Dodge, K. A. (1980). Social cognition and children's aggressive behavior. *Child Development*, *51*, 162-170.

Dodge, K. A. (1991). The structure and function of reactive and proactive aggression. In D. J. Pepler & K. H. Rubin (Eds.), *The Development and treatment of childhood aggression* (pp. 201–218). Hillsdale: Lawrence Erlbaum Associates, Inc.

Dodge, K. A. (2006). Translational science in action: Hostile attributional style and the development of aggressive behavior problems. *Development and Psychopathology, 18*, 791-814.

Dodge, K. A., & Coie, J. D. (1987). Social-information-processing factors in reactive and proactive aggression in children's peer groups. *Journal of Personality and Social Psychology*, *53*, 1146-1158.

Dodge, K. A., Murphy, R. R., & Buchsbaum, K. (1984). The assessment of intention-cue detection skills in children: Implications for developmental psychopathology. *Child Development*, *55,* 163-173.

Dodge, K. A., & Price, J. M. (1994). On the relation between social information processing and socially competent behavior in early school-aged children. *Child Development*, *65*, 1385-1397.

Dumais, A., Lesage, A. D., Alda, M., Rouleau, G., Dumont, M., Chawky, N., & Turecki, G. (2005). Risk factors for suicide completion in major depression: a case-control study of impulsive and aggressive behaviors in men. *American Journal of Psychiatry*, *162*, 2116-2124.

Eckhardt, C. I., & Cohen, D. J. (1997). Attention to anger-relevant and irrelevant stimuli following naturalistic insult. *Personality and Individual Differences*, *23*, 619-629.

Farrar, K., & Krcmar, M. (2006). Measuring state and trait aggression: A short, cautionary tale. *Media Psychology, 8*, 127–138.

Ferrari, G. R., Möbius, M., van Opdorp, A., Becker, E. S., & Rinck, M. (2016). Can't look away: An eye-tracking based attentional disengagement training for depression. *Cognitive Therapy and Research*, *40*, 672-686.

Foulsham, T., Gray, A., Nasiopoulos, E., & Kingstone, A. (2013). Leftward biases in picture scanning and line bisection: A gaze-contingent window study. *Vision Research*, *78*, 14-25.

Gerdes, A. B., Alpers, G. W., & Pauli, P. (2008). When spiders appear suddenly: Spider-phobic patients are distracted by task-irrelevant spiders. *Behaviour Research and Therapy*, *46*, 174-187.

Glaholt, M. G., & Reingold, E. M. (2011). Eye movement monitoring as a process tracing methodology in decision making research. *Journal of Neuroscience, Psychology, and Economics*, *4*, 125-146.

Hallion, L. S., & Ruscio, A. M. (2011). A meta-analysis of the effect of cognitive bias modification on anxiety and depression. *Psychological Bulletin*, *137*, 940-958.

Haselton, M. G., Nettle, D., Murray, D. R. (2015). The evolution of cognitive bias. In D. M. Buss (Ed.), *The Handbook of evolutionary psychology* (2nd ed., pp. 968-987). Hoboken, NJ: Wiley.

Hawkins, K. A., & Cougle, J. R. (2013). Effects of interpretation training on hostile attribution bias and reactivity to interpersonal insult. *Behavior Therapy, 44*, 479-488.

Heeren, A., Mogoașe, C., Philippot, P., & McNally, R. J. (2015). Attention bias modification for social anxiety: A systematic review and meta-analysis. *Clinical Psychology Review*, *40*, 76-90.

Hiemstra, W., de Castro, B. O., & Thomaes, S. (2018). Reducing aggressive children's hostile attributions: A Cognitive Bias Modification Procedure. *Cognitive Therapy and Research*, *43,* 387-398.

Hirsch, C. R., & Clark, D. M. (2004). Information-processing bias in social phobia. *Clinical Psychology Review*, *24*, 799-825.

Hirsch, C. R., Clark, D. M., & Mathews, A. (2006). Imagery and interpretations in social phobia: Support for the combined cognitive biases hypothesis. *Behavior Therapy*, *37*, 223-236.

Holmes, E. A., Lang, T. J., & Shah, D. M. (2009). Developing interpretation bias modification as a "cognitive vaccine" for depressed mood: Imagining positive events makes you feel better than thinking about them verbally. *Journal of Abnormal Psychology*, *118*, 76-88.

Holmes, E. A., Mathews, A., Dalgleish, T., & Mackintosh, B. (2006). Positive interpretation training: Effects of mental imagery versus verbal training on positive mood. *Behavior Therapy*, *37*, 237-247.

Horsley, T. A., de Castro, B. O., & Van der Schoot, M. (2010). In the eye of the beholder: Eye-tracking assessment of social information processing in aggressive behavior. *Journal of Abnormal Child Psychology, 38*, 587-599.

Huesmann, L. R., & Guerra, N. G. (1997). Children's normative beliefs about aggression and aggressive behavior. *Journal of Personality and Social Psychology*, *72*, 408-419.

Huijding, J., Mayer, B., Koster, E. H., & Muris, P. (2011). To look or not to look: An eye movement study of hypervigilance during change detection in high and low spider fearful students. *Emotion*, *11*, 666-674.

Hymel, S., Comfort, C., Schonert-Reichl, K., & McDougall, P. (1996). Academic failure and school dropout: The influence of peers. In J. Juvonen & K. R.Wentzel (Eds.), *Social motivation: Understanding children's school adjustment,* (pp. 313-345). New York: Cambridge University Press.

Ialongo, N. S., Vaden-Kieman, N., & Kellam, S. (1998). Early peer rejection and aggression: Longitudinal relations with adolescent behavior. *Journal of Developmental and Physical Disabilities*, *10*, 199-213.

Klein Tuente, S., Bogaerts, S., & Veling, W. (2019). Hostile attribution bias and aggression in adults-a systematic review. *Aggression and Violent Behavior*, *46,* 66-81.

Krug, E. G., Mercy, J. A., Dahlberg, L. L., & Zwi, A. B. (2002). The world report on violence and health. *The Lancet*, *360*, 1083-1088.

Kulasegaram, K., Min, C., Howey, E., Neville, A., Woods, N., Dore, K., & Norman, G. (2015). The mediating effect of context variation in mixed practice for transfer of basic science. *Advances in Health Sciences Education*, *20*, 953-968.

Laue, C., Griffey, M., Lin, P. I., Wallace, K., Van der Schoot, M., Horn, P., Pedapati, E., & Barzman, D. (2018). Eye gaze patterns associated with aggressive tendencies in adolescence. *Psychiatric Quarterly*, *89*, 747-756.

Lazarov, A., Pine, D. S., & Bar-Haim, Y. (2017). Gaze-contingent music reward therapy for social anxiety disorder: A randomized controlled trial. *American Journal of Psychiatry*, *174*, 649-656.

Lee, A. H., & DiGiuseppe, R. (2018). Anger and aggression treatments: a review of meta-analyses. *Current Opinion in Psychology*, *19*, 65-74.

Lin, P. I., Hsieh, C. D., Juan, C. H., Hossain, M. M., Erickson, C. A., Lee, Y. H., & Su, M. C. (2016). Predicting aggressive tendencies by visual attention bias associated with hostile emotions. *PloS One*, *11*, 1-8.

Lisk, S. C., Pile, V., Haller, S. P., Kumari, V., & Lau, J. Y. (2018). Multisession cognitive bias modification targeting multiple biases in adolescents with elevated social anxiety. *Cognitive Therapy and Research*, *42*, 581-597.

Lothmann, C., Holmes, E. A., Chan, S. W., & Lau, J. Y. (2011). Cognitive bias modification training in adolescents: effects on interpretation biases and mood. *Journal of Child Psychology and Psychiatry*, *52*, 24-32.

MacLeod, C., & Clarke, P. J. (2015). The attentional bias modification approach to anxiety intervention. *Clinical Psychological Science*, *3*, 58-78.

MacLeod, C., & Mathews, A. (2012). Cognitive bias modification approaches to anxiety. *Annual Review of Clinical Psychology*, *8*, 189-217.

MacLeod, C., Rutherford, E., Campbell, L., Ebsworthy, G., & Holker, L. (2002). Selective attention and emotional vulnerability: Assessing the causal basis of their association through the experimental manipulation of attentional bias. *Journal of Abnormal Psychology*, *111*, 107-123.

Mathews, A., & Mackintosh, B. (2000). Induced emotional interpretation bias and anxiety. *Journal of Abnormal Psychology*, *109*, 602-615.

Maoz, K., Adler, A. B., Bliese, P. D., Sipos, M. L., Quartana, P. J., & Bar-Haim, Y. (2017). Attention and interpretation processes and trait anger experience, expression, and control. *Cognition and Emotion*, *31*, 1453-1464.

McGowan, N., Sharpe, L., Refshauge, K., & Nicholas, M. K. (2009). The effect of attentional re-training and threat expectancy in response to acute pain. *Pain*, *142*, 101-107.

Mobini, S., Mackintosh, B., Illingworth, J., Gega, L., Langdon, P., & Hoppitt, L. (2014). Effects of standard and explicit cognitive bias modification and computer-administered cognitive-behaviour therapy on cognitive biases and social anxiety. *Journal of Behavior Therapy and Experimental Psychiatry*, *45*, 272-279.

Naim, R., Kivity, Y., Bar-Haim, Y., & Huppert, J. D. (2018). Attention and interpretation bias modification treatment for social anxiety disorder: A randomized clinical trial of efficacy and synergy. *Journal of Behavior Therapy and Experimental Psychiatry*, *59*, 19-30.

Novaco, R.W. (1994). Anger as a risk factor for violence among the mentally disordered. In J. Monahan & H. Steadman (Eds.), *Violence and mental disorder: Developments in risk assessment* (pp. 21-59). Chicago: University of Chicago Press.

Panagiotidis, D., Clemens, B., Habel, U., Schneider, F., Schneider, I., Wagels, L., & Votinov, M. (2017). Exogenous testosterone in a non-social provocation paradigm potentiates anger but not behavioral aggression. *European Neuropsychopharmacology*, *27*, 1172-1184.

Penton-Voak, I. S., Thomas, J., Gage, S. H., McMurran, M., McDonald, S., & Munafò, M. R. (2013). Increasing recognition of happiness in ambiguous facial expressions reduces anger and aggressive behavior. *Psychological Science, 24*, 688-697.

Poulin, F., & Boivin, M. (2000). Reactive and proactive aggression: Evidence of a two-factor model. *Psychological Assessment*, *12*, 115-122.

Price, R. B., Greven, I. M., Siegle, G. J., Koster, E. H., & De Raedt, R. (2016). A novel attention training paradigm based on operant conditioning of eye gaze: Preliminary findings. *Emotion*, *16*, 110-116.

Psychology Software Tools Inc. 2002. E-Prime (Version 2.0) [computer software]. Psychology Software Tolls Pittsburg, PA.

Quiggle, N. L., Garber, J., Panak, W. F., & Dodge, K. A. (1992). Social information processing in aggressive and depressed children. *Child Development, 63*, 1305–1320.

Raine, A., Dodge, K., Loeber, R., Gatzke-Kopp, L., Lynam, D., Reynolds, C., Stouthamer-Loeber, M., & Liu, J. (2006). The reactive-proactive aggression questionnaire: Differential correlates of reactive and proactive aggression in adolescent boys. *Aggressive Behavior, 32*, 159-171.

Sanchez-Lopez, A., De Raedt, R., van Put, J., & Koster, E. H. (2019). A novel process-based approach to improve resilience: Effects of computerized mouse-based (gaze) contingent attention training (MCAT) on reappraisal and rumination. *Behaviour Research and Therapy*, *118*, 110-120.

Sanchez, A., Everaert, J., & Koster, E. H. (2016). Attention training through gaze-contingent feedback: Effects on reappraisal and negative emotions. *Emotion*, *16*, 1074-1085.

Schippell, P. L., Vasey, M. W., Cravens-Brown, L. M., & Bretveld, R. A. (2003). Suppressed attention to rejection, ridicule, and failure cues: A unique correlate of reactive but not proactive aggression in youth. *Journal of Clinical Child and Adolescent Psychology*, *32*, 40-55.

Schoenmakers, T. M., de Bruin, M., Lux, I. F., Goertz, A. G., Van Kerkhof, D. H., & Wiers, R. W. (2010). Clinical effectiveness of attentional bias modification training in abstinent alcoholic patients. *Drug and Alcohol Dependence*, *109*, 30-36.

Schönenberg, M., & Jusyte, A. (2014). Investigation of the hostile attribution bias toward ambiguous facial cues in antisocial violent offenders. *European Archives of Psychiatry and Clinical Neuroscience*, *264*, 61-69.

Simons, R. L., & Burt, C. H. (2011). Learning to be bad: Adverse social conditions, social schemas, and crime. *Criminology*, *49*, 553-598.

Skara, S., Pokhrel, P., Weiner, M. D., Sun, P., Dent, C. W., & Sussman, S. (2008). Physical and relational aggression as predictors of drug use: Gender differences among high school students. *Addictive Behaviors*, *33*, 1507-1515.

Smeets, E., Jansen, A., & Roefs, A. (2011). Bias for the (un)attractive self: On the role of attention in causing body (dis)satisfaction. *Health Psychology*, *30*, 360-367.

Smeijers, D., Rinck, M., Bulten, E., van den Heuvel, T., & Verkes, R. J. (2017). Generalized hostile interpretation bias regarding facial expressions: Characteristic of pathological aggressive behavior. *Aggressive Behavior*, *43*, 386-397.

Smith, P., & Waterman, M. (2003). Processing bias for aggression words in forensic and nonforensic samples. *Cognition & Emotion*, *17*, 681-701.

Smith, P., & Waterman, M. (2004). Role of experience in processing bias for aggressive words in forensic and non-forensic populations. *Aggressive Behavior*, *30*, 105-122.

Speilberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the state-trait anxiety inventory* (STAI). San Diego, CA: Mindgarden.

Swogger, M. T., Walsh, Z., Christie, M., Priddy, B. M., & Conner, K. R. (2015). Impulsive versus premeditated aggression in the prediction of violent criminal recidivism. *Aggressive Behavior*, *41*, 346-352.

Taylor, S. P. (1967). Aggressive behavior and physiological arousal as a function of provocation and the tendency to inhibit aggression1. *Journal of Personality, 35*, 297-310.

Troop-Gordon, W., Gordon, R. D., Vogel-Ciernia, L., Ewing Lee, E., & Visconti, K. J. (2018). Visual attention to dynamic scenes of ambiguous provocation and children's aggressive behavior. *Journal of Clinical Child & Adolescent Psychology*, *47*, 925-940.

Van Bockstaele, B., Verschuere, B., Tibboel, H., De Houwer, J., Crombez, G., & Koster, E. H. (2013). A review of current evidence for the causal impact of attentional bias on fear and anxiety. *Psychological Bulletin, 140*, 682-721.

Van Honk, J., Tuiten, A., de Haan, E., vann de Hout, M., & Stam, H. (2001a). Attentional biases for angry faces: Relationships to trait anger and anxiety. *Cognition & Emotion*, *15*, 279-297.

Van Honk, J., Tuiten, A., van den Hout, M., Putman, P., de Haan, E., & Stam, H. (2001b). Selective attention to unmasked and masked threatening words: Relationships to trait anger and anxiety. *Personality and Individual Differences*, *30*, 711-720.

Vassilopoulos, S. P., Brouzos, A., & Andreou, E. (2015). A multi-session attribution modification program for children with aggressive behavior: Changes in attributions, Anger score estimates, and self-reported aggression. *Behavioral and Cognitive Psychotherapy*, *43*, 538-548.

Waas, G. A. (1988). Social attributional biases of peer-rejected and aggressive children. *Child Development, 59*, 969-975.

Wadlinger, H. A., & Isaacowitz, D. M. (2008). Looking happy: the experimental manipulation of a positive visual attention bias. *Emotion*, *8*, 121-126.

Wang, Q., Celebi, F. M., Flink, L., Greco, G., Wall, C., Prince, E., Lansiquot, S., Chawarska, K., Kim, S. E., Boccanfuso, L., DiNicola. L., & Shic. F. (2015). Interactive eye tracking for gaze strategy modification. In *Proceedings of IDC 2015: The 14th International Conference on Interaction Design and Children* (pp. 247-250). Association for Computing Machinery, Inc. https://doi.org/10.1145/2771839.2771888

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*, 1063-1070.

Werner, N. E., & Nixon, C. L. (2005). Normative beliefs and relational aggression: An investigation of the cognitive bases of adolescent aggressive behavior. *Journal of Youth and Adolescence*, *34*, 229-243.

White, L. K., Suway, J. G., Pine, D. S., Bar-Haim, Y., & Fox, N. A. (2011). Cascading effects: The influence of attention bias to threat on the interpretation of ambiguous information. *Behaviour Research and Therapy*, *49*, 244-251.

Wilkowski, B. M., & Robinson, M. D. (2008). The cognitive basis of trait anger and reactive aggression: An integrative analysis. *Personality and Social Psychology Review*, *12*, 3-21.

Wilkowski, B. M., Robinson, M. D., Gordon, R. D., & Troop-Gordon, W. (2007). Tracking the evil eye: Trait anger and selective attention within ambiguously hostile scenes. *Journal of Research in Personality, 41*, 650-666.

World Medical Association. (2001). Declaration of Helsinki World Medical Association Declaration of Helsinki. *Bulletin of the World Health Organization*, *79*, 373-374.

# S

## Nederlandse Samenvatting

Vooruitgang in het begrijpen van de rol van cognitieve biases (vertekeningen) van aandacht en interpretatie bij agressie hebben geleid tot een belangrijke nieuwe interesse in het toepassen van deze kennis op interventieonderzoek (de Castro, Veerman, Koops, Joop, & Monshouwer, 2002; Troop-Gordon, Gordon, Vogel-Ciernia, Lee, & Visconti, 2018). De resultaten van de eerste interventiestudies hebben aangetoond dat het modificeren van een specifieke informatieverwerkingsbias met behulp van het cognitieve bias modificatie (CBM) paradigma kan leiden tot een verandering van de beoogde bias en een significante vermindering van woede en agressie (e.g., Hawkins & Cougle, 2013; Vassilopoulos, Brouzos & Andreou, 2015). Aan het begin van dit proefschrift waren de agressiestudies echter beperkt in hun onderzoek naar de effecten van de CBM-paradigma's als het gaat om de interpretaties (CBM-I). Bovendien zijn er, in de context van agressie, geen onderzoeken naar de effectiviteit van cognitieve bias modificatie paradigma's gericht op aandacht (CBM-A) voor zowel bias als symptoomreductie.

Daarnaast worden in agressiestudies de interpretatiebiases meestal afzonderlijk beoordeeld en getraind met behulp van vignetten die hypothetische provocerende sociale situaties beschrijven waarin de intentie van de dader ambigue is (e.g., Hawkins & Cougle, 2013; Vassilopoulos et al., 2015). In werkelijkheid zijn visuele non-verbale signalen zoals gezichts- en lichaamsuitdrukkingen ook belangrijke signalen die wat zeggen over de interne toestand van anderen en hun intenties (Cadesky, Mota, & Schachar, 2000). Dit doet de vraag rijzen of het meenemen van de visuele ambigue sociale situaties in het trainingsproces een meer informatierijke en naturalistische context zou kunnen bieden, wat het effect van het trainingsproces vergroot op zowel de interpretatiebias als de aandachtsbias. Het doel van dit proefschrift is om te onderzoeken of een nieuwe CBM-procedure met behulp van picturale stimuli kan worden gebruikt om de aandachts- en interpretatiebiases te modificeren in de context van agressie en het onderzoeken van het effect van deze modificaties op agressie.

Verder suggereert de cognitieve bias hypothese dat cognitieve biases met elkaar samenhangen en dat trainingsprocedures die gericht zijn op een combinatie van biases een grotere impact hebben op stoornissen dan trainingsprocedures die gericht zijn op een afzonderlijk cognitief proces (Hirsch, Clark & Mathews, 2006). Dit roept de vraag op hoe cognitieve biases van aandacht en interpretatie op elkaar kunnen inwerken en kunnen bijdragen tot agressief gedrag. Het roept ook de vraag op of er een toegevoegde waarde is van een gecombineerde bias training gericht op

zowel aandachts- als interpretatiebias ten opzichte van een enkele bias training in zowel beide biases als agressiereductie.

Om te onderzoeken of het mogelijk is om agressie te modificeren door het aanpassen van de aandachts- en interpretatiebiases en het gebruik van picturale stimuli te bestuderen om meer trainingseffecten te bewerkstelligen, richt dit onderzoek zich op de volgende vier vragen: (1) Kan een nieuwe CBM training procedure met behulp van picturale stimuli gebruikt worden om de interpretatie- en aandachtsbiases te veranderen in de context van agressie? (2) Leiden veranderingen in de aandachts- of interpretatiebiases tot veranderingen in agressie? (3) Hoe interacteren de aandachts- en interpretatiebiases bij het in stand houden van agressie? en (4) Is een gecombineerde bias CBM training procedure effectiever dan een enkele bias CBM training procedure op zowel beide biases als agressiereductie? De bevindingen van de vier onderzoeksvragen worden hieronder samengevat.



**Kan een nieuwe CBM training procedure met behulp van picturale stimuli gebruikt worden om de interpretatie- en aandachtsbiases te veranderen in de context van agressie?**

In de pilotstudie, beschreven in hoofdstuk 2, is onderzocht of een nieuwe CBM-I procedure met behulp van picturale stimuli de interpretaties beïnvloedt. De resultaten toonden aan dat een enkele sessie van positieve interpretatietraining met behulp van picturale stimuli de pro-sociale interpretatiebias verhoogde. De negatieve training leidde echter niet tot een toename van de vijandige interpretatiebias.

Een van de belangrijkste beperkingen van dit onderzoek was het ontbreken van een controlegroep. Hierdoor konden we niet concluderen dat de positieve verandering in interpretatiebiases volledig te wijten was aan onze CBM-I training procedure. Om dit probleem aan te pakken, is er een controleconditie opgenomen in de studies die beschreven staan in hoofdstuk 4 en 5. De resultaten van het in hoofdstuk 4 beschreven onderzoek laten zien dat de training de interpretatiebias in een pro-sociale richting heeft veranderd ten opzichte van de controletraining. Daarnaast heeft de negatieve CBM-I training de interpretatiebiases in een vijandige richting veranderd, maar niet meer dan in de controletraining. De bevindingen beschreven in hoofdstuk 5 toonden aan dat de pro-sociale interpretatiebiases van de deelnemers toenamen van pre- tot posttraining. Vanwege methodologische problemen met betrekking tot de controleconditie (beschreven in hoofdstuk 5), waren de veranderingen in de pro-sociale interpretatiebias echter niet anders dan in de controleconditie. Desalniettemin zijn de effecten van

onze CBM-I op de pro-sociale interpretatiebias consistent in alle drie de studies. Dit lijkt de conclusie te rechtvaardigen dat onze trainingsprocedure met behulp van picturale stimuli met succes kan leiden tot meer pro-sociale interpretatie verwerkingsstijlen in de context van agressie.

In de studie beschreven in hoofdstuk 3 is onderzocht of onze nieuwe CBM-procedure met behulp van picturale stimuli invloed heeft op de aandachtsbias. De resultaten geven aan dat de *gaze-contingent* CBM-A training in de positieve conditie resulteerde in een toename van de aandacht voor adaptieve signalen (d.w.z. het gezicht van de dader) dan voor maladaptieve signalen (d.w.z. de negatieve uitkomst van de situatie) in een ambigue sociale situatie.

De huidige bevindingen suggereren dat onze nieuwe CBM training procedures de aandachtbias (hoofdstuk 3 en 5) en de interpretatiebias (hoofdstuk 2, 4 en 5) op een pro-sociale manier kunnen veranderen.

**Leiden veranderingen in de aandachts- of interpretatiebiases tot veranderingen in agressie?**

De resultaten van de pilotstudie, beschreven in hoofdstuk 2, waren veelbelovend. De CBM-I training resulteerde niet alleen in een toename van de pro-sociale interpretatiebias, maar leidde ook tot een afname van verbale agressie en zelfgerapporteerde woede en een toename van het zelfgerapporteerde geluk. Bovendien bleek in de negatieve training dat hoe meer de deelnemers hun interpretaties in een vijandige manier veranderden, hoe agressiever zij reageerden op het *Taylor Aggression Paradigm* (TAP; Taylor, 1967). Helaas waren we niet in staat om deze resultaten te reproduceren in de studies beschreven in hoofdstuk 4 en 5, waarin de verandering in interpretatiebias geen invloed had in de verwachte richting op agressie of stemming.

In de studie beschreven in hoofdstuk 3 is onderzocht wat het effect van een *gaze-contingent* CBM-A training op agressie is. De resultaten toonden aan dat zowel de positieve als negatieve trainingscondities niet leidden tot veranderingen in de verwachte richting op agressie en stemming.

Samenvattend zijn de effecten van de CBM-I op agressie in het huidige proefschrift inconsistent en vonden we geen bewijs voor de effectiviteit van CBM-A op agressie. Om deze reden zijn we niet in staat om harde conclusies te trekken over de effectiviteit van ons CBM-training paradigma op agressie.

**Hoe interacteren de aandacht en interpretatiebiases bij het in stand houden van agressie?**

De studie beschreven in hoofdstuk 3 richtte zich op de vraag of het teweegbrengen van veranderingen in de aandachtsbias door gebruik te maken van CBM-A invloed zou hebben op de interpretatie van de daaropvolgende ambigue sociale informatie. De resultaten toonden aan dat de

aandachtstraining niet leidde tot veranderingen in de interpretatiebias in de verwachte richting. Dat wil zeggen dat in beide trainingsgroepen de interpretatiebias pro-sociaal toenam en dus onafhankelijk leek van de geïnduceerde aandachtsbias.

In de studie die beschreven staat in hoofdstuk 4 hebben we het effect in de tegenovergestelde richting onderzocht en de impact van de training interpretatiebias getest door gebruik te maken van CBM-I op de aandachtsbias.

De resultaten toonden aan dat in alle trainingscondities de aandachtsbias in een pro-sociale richting veranderde. In de studie die beschreven staat in hoofdstuk 5 hebben we opnieuw de relatie tussen de interpretatiebias en aandachtsbias aan de orde gesteld door middel van een CBM-I training procedure in een van de training condities. De resultaten toonden aan dat CBM-I leidde tot een toename van zowel de pro-sociale interpretatie als de aandachtsbias van pre- tot posttraining. In tegenstelling tot onze verwachtingen waren die veranderingen niet anders dan in de controleconditie.



Het doel van de studie die beschreven staat in hoofdstuk 4 was om de bevindingen van hoofdstuk 2 uit te breiden met de vraag of de CBM-I training ook invloed had op de interpretatie van gezichtsuitdrukkingen door de deelnemers. De resultaten van de studie toonden aan dat als gevolg van het beperkte effect van de CBM-I op de interpretatiebias van de intentie, het transfereffect van de gemodificeerde interpretatie van de intentie naar de interpretatie van de gezichtsuitdrukking beperkt kan zijn geweest. Daarom wilden we de relatie tussen de interpretatie van de intentie en de interpretatie van de gezichtsuitdrukking verder onderzoeken.

Zo hebben we in het onderzoek dat in hoofdstuk 5 beschreven staat opnieuw de interpretatie van gezichtsuitdrukkingen gemeten. Interessant genoeg blijkt uit de resultaten dat in alle trainingscondities niet alleen de aandachtsbias en interpretatiebias in een pro-sociale richting veranderden, maar ook de pro-sociale interpretatie van gezichtsuitdrukkingen in een pro-sociale richting veranderde van pre- naar posttraining. Onze bevindingen suggereren dat een modificatie van de interpretatiebias van de intentie een effect kan hebben op de beïnvloeding van andere (verwante) vormen van biases. Echter is het in dit stadium slechts speculeren en is het nodig om de associatie tussen de biases over de intenties en de interpretatie van gezichtsuitdrukkingen in de toekomst verder te onderzoeken.

Samenvattend suggereren de huidige bevindingen dat de biases in interpretatie en aandacht kunnen interacteren in de context van agressie. In het bijzonder geeft dit proefschrift de eerste

resultaten die aangeven dat de manier waarop individuen de intenties van anderen interpreteren van invloed is op hun aandacht voor sociale signalen en hoe individuen ambigue gelaatsuitdrukkingen interpreteren.

**Is een gecombineerde bias CBM training procedure effectiever dan een enkele bias CBM training procedure op zowel beide biases als agressiereductie?**

In de studie die in hoofdstuk 5 staat beschreven, is de toegevoegde waarde van een gecombineerd CBM gericht op zowel de aandachtsbias als interpretatiebias (CBM-AI) ten opzichte van één enkele bias training (bv. CBM-I) bestudeerd. De resultaten van de studie toonden aan dat een enkele sessie van CBM-AI de aandacht voor adaptieve signalen en pro-sociale interpretatie van de intentie verhoogde. In tegenstelling tot onze verwachtingen hebben we echter vastgesteld dat deze veranderingen niet significant verschillen van de veranderingen in de CBM-I en de controlecondities. De resultaten beschreven in hoofdstuk 5 toonden ook aan dat geen van de trainingscondities een effect had op agressiemaatregelen.

Samenvattend hebben we geen bewijs om harde conclusies te trekken over de vraag of een gecombineerde bias CBM training procedure effectiever is dan een enkele bias CBM training procedure in zowel beide biases als agressiereductie.

**Conclusie**

Op basis van het SIP-model (Crick & Dodge, 1994) is in dit proefschrift onderzocht of beide nieuwe CBM-procedures (CBM-I met behulp van picturale stimuli om de interpretatiebias te modificeren en CBM-A met behulp van picturale stimuli om de aandachtbias aan te passen) zouden kunnen leiden tot een verandering die gericht is op de cognitieve bias en agressie. Op basis van de resultaten die beschreven staan in dit proefschrift kunnen een aantal conclusies worden getrokken. Ten eerste, onze bevindingen tonen aan dat elke CBM succesvol was in het modificeren van de gerichte bias op een pro-sociale manier. Ten tweede, onze CBM training procedures waren niet effectief in het verminderen van agressie of negatieve stemming. Ten derde konden we geen bewijs vinden van de onderlinge relatie tussen de aandachtsbias en interpretatiebias in de context van agressie. Sommige van onze bevindingen kunnen echter suggereren dat, in vergelijking met de aandachtsbiases, een interpretatiebias een meer generaliseerbaar effect heeft in de context van agressie. Tot slot vonden we dat een gecombineerde bias training geen extra effectiviteit heeft ten opzichte van een enkele bias training in agressiereductie. Samenvattend voegt dit proefschrift kennis toe gerelateerd aan de rol van de aandachtsbias en interpretatiebias in de context van

agressie. Maar er blijven wel een aantal kwesties onduidelijk en daarom is er meer onderzoek nodig. We hopen dat de studies die beschreven staan in dit proefschrift inspireren om verder onderzoek uit te voeren naar CBM-training paradigma's die kunnen helpen bij het vinden van de meest effectieve manier om positieve veranderingen in de cognitieve informatieverwerking, die ten grondslag ligt aan agressie, te bewerkstelligen.

# C

## Curriculum Vitae

## Curriculum Vitae

Nouran AlMoghrabi was born on the 6$^{Th}$ of April 1986, in Dammam, Saudi Arabia. She obtained a Bachelor's Degree in Educational Psychology from the University of Bahrain in 2009. Following this, she went on to complete her postgraduate studies in Clinical Psychology at Bangor University, United Kingdom, and was awarded a Master's of Science degree in 2011. Subsequently, she was appointed as a lecturer at Princess Nourah bint Abdulrahman University in Riyadh, Saudi Arabia, where she was also granted funding to pursue a PhD in the Netherlands. She started working in the capacity of a PhD candidate from September 2014 in the Department of Psychology, Education and Child Studies at Erasmus University Rotterdam. Her PhD project focused on aggression-related biases of attention and interpretation and its modification using computerized Cognitive Bias Modification training paradigms. After the conclusion of her PhD, she will be moving back to Saudi Arabia to start working as an assistant professor in the Department of Psychology at Princess Nourah bint Abdulrahman University.



## Publications

AlMoghrabi, N., Huijding, J., & Franken, I. H. (2018). The effects of a novel hostile interpretation bias modification paradigm on hostile interpretations, mood, and aggressive behavior. *Journal of Behavior Therapy and Experimental Psychiatry*, *58*, 36-42.

AlMoghrabi, N., Huijding, J., Mayer, B., & Franken, I. H. (2019). Gaze-contingent attention bias modification training and its effect on attention, interpretations, mood, and aggressive behavior. *Cognitive Therapy and Research, 43*, 861-873.

## Submitted manuscripts

AlMoghrabi, N., Franken, I. H., Mayer, B., van der Schoot, M., & Huijding, J. CBM-I training and its effect on interpretations of intent, facial expressions, attention and aggressive behavior.

AlMoghrabi, N., Franken, I. H., Mayer, B., & Huijding, J. A single-session combined cognitive bias modification training targeting attention and interpretation biases in aggression.

## Conference presentations

AlMoghrabi, N., Huijding, J., Mayer, B., & Franken, I. H. (2018). Gaze-contingent attention bias modification training and its effect on attention, interpretations, mood, and aggressive behavior. Poster presented at 23$^{rd}$ International Society for Research on Aggression World Meeting (ISRA), Paris.

AlMoghrabi, N., Huijding, J., & Franken, I. H. (2017). The effects of a novel hostile interpretation bias modification paradigm on hostile interpretations, mood, and aggressive

behavior. Presentation at Graduate Research Day (GRD) at Erasmus University Rotterdam, Rotterdam.

AlMoghrabi, N., Huijding, J., & Franken, I. H. (2016). The effects of a novel hostile interpretation bias modification paradigm on hostile interpretations, mood, and aggressive behavior. Poster presented at 22nd International Society for Research on Aggression World Meeting (ISRA), Sydney.