# Modification of hostile attribution bias reduces self-reported reactive aggressive behaviour in adolescents

Van Bockstaele, B.; van der Molen, M.J.; van Nieuwenhuijzen, M.; Salemink, E.

**DOI**
10.1016/j.jecp.2020.104811

**Publication date**
2020

**Document Version**
Final published version

**Published in**
Journal of Experimental Child Psychology

Journal of Experimental Child Psychology 194 (2020) 104811

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

Brief Report

# Modification of hostile attribution bias reduces self-reported reactive aggressive behavior in adolescents

Bram Van Bockstaele [a,b,*], Mariët J. van der Molen [c], Maroesjka van Nieuwenhuijzen [d], Elske Salemink [e]

[a] Department of Psychology, University of Amsterdam, 1018 WT Amsterdam, the Netherlands
[b] School of Psychological Science, University of Western Australia, WA, 6009 Crawley, Australia
[c] Mariët van der Molen Onderzoek, Onderwijs & Advies, 1086 ZV Amsterdam, the Netherlands
[d] Expertise Centre William Schrikker, P.O. Box 12685, 1100 AR Amsterdam, the Netherlands
[e] Department of Clinical Psychology, Utrecht University, P.O. Box 80140, 3508 TC Utrecht, the Netherlands

ARTICLE INFO

*Article history:*
Received 28 June 2019
Revised 22 October 2019
Available online 21 February 2020

*Keywords:*
Hostile attribution bias
Aggression
Interpretation bias
Cognitive bias modification

ABSTRACT

Aggressive individuals more readily interpret others' motives and intentions in ambiguous situations as hostile. This hostile attribution bias has been argued to be causally involved in the development and maintenance of aggression, making it a target for interventions. In our current study, adolescents selected for high levels of aggression ($N$ = 39) were assigned to either a test–retest control group or a five-session hostile attribution bias modification training, in which they were trained to make more benign interpretations of ambiguously provocative social situations. Before and after the training, we assessed hostile attribution bias and both reactive and proactive self-reported aggression in both groups. The training not only tended to produce the expected reduction in hostile attribution bias but also crucially led to decreased levels of reactive but not proactive aggression compared with the control group. Our results thus support the idea that hostile attribution bias can be targeted using training techniques and that such training-induced changes in bias may reduce aggression. However, future studies using an active control group and multiple outcome measures are needed to address the long-term effects of training.



* Corresponding author at: School of Psychological Science, University of Western Australia, WA, 6009 Crawley, Australia.
*E-mail address:* Bram.Vanbockstaele@uwa.edu.au (B. Van Bockstaele).

https://doi.org/10.1016/j.jecp.2020.104811
0022-0965/

## Introduction

According to Dodge's (2006) model of social information processing, aggressive individuals are more likely to interpret others' motives in ambiguous social events as provocative rather than harmless or accidental. This tendency is termed hostile attribution bias (Nasby, Hayden, & Depaulo, 1980) and is a robust empirical finding (Orobio de Castro, Veerman, Koops, Bosch, & Monshouwer, 2002) considered to causally contribute to the development and maintenance of aggressive behaviors: interpreting others' motives as hostile increases the likelihood that one will react aggressively. Some studies have indeed shown that interventions reducing hostile attribution bias also reduce aggression (e.g., Hudley & Graham, 1993). However, these interventions were typically both time and cost intensive, and they resulted only in relatively modest effects on aggression (Dodge, 2006).

More recently, studies have attempted to directly change hostile attribution bias and subsequent aggression using more implicit training procedures based on the interpretation bias modification literature. These studies originated from the anxiety field and are based on the idea that anxious individuals interpret ambiguous stimuli or situations as negative or threatening. This negative interpretation bias has been causally related to anxiety, with training-induced reductions in negative interpretation bias leading to reductions in anxiety (Krebs et al., 2018). A similar approach has also been adopted in studies on hostile attribution bias and aggressive behaviors. Penton-Voak et al. (2013) asked participants to classify pictures of morphed emotionally ambiguous faces varying on the continuum between happy and angry. In a facial emotion recognition training group, they progressively reinforced participants to classify more ambiguously angry faces as happy. The training resulted not only in a more positive classification of ambiguous faces (i.e., participants in the training group were more likely to classify relatively angry ambiguous faces as happy) but also in decreases in self-reported aggression and anger in healthy adults (Penton-Voak et al., 2013, Experiments 1 and 3) and decreases in both staff-rated and self-reported aggressive behavior in high-risk adolescents (Experiment 2). Promising as these findings are, Hiemstra, Orobio de Castro, and Thomaes (2019) could not replicate them in two experiments. Using a similar procedure to train clinically referred aggressive boys to classify ambiguous faces as happy, they found that training reduced the hostile interpretation of ambiguous faces, but there were no training effects on measures of aggression.

Next to the facial emotion recognition training, researchers have also used the ambiguous scenario training (Mathews & Mackintosh, 2000). In this paradigm, participants read emotionally ambiguous scenarios. In the last sentence, a disambiguating word is presented as a word fragment, and participants are required to complete the word fragment, thereby disambiguating the scenario in a positive or negative manner. Using this paradigm, Hawkins and Cougle (2013) trained undergraduate students in a single lab session to make either benign or hostile interpretations of others' ambiguous intentions. The benign training not only affected participants' tendency to interpret others' intentions as more friendly than the hostile training but also had a marked effect on anger reactivity: the benign training group reported less anger and showed less irritation in response to a staged insult than the hostile training group. Vassilopoulos, Brouzos, and Andreou (2015) used a similar training in a sample of children displaying high levels of aggressive behavior. Children who completed a three-session benign attribution bias training showed increased benign attributions and decreased hostile attributions and self-reported aggression following the training, whereas all measures remained stable in a test–retest control group.

With our current study, we aimed to specify and strengthen the effects of training-induced changes in hostile attribution bias on aggression. First, hostile attribution bias has been related especially to reactive aggression (i.e., aggressive reactions to perceived threats or frustrations) rather than proactive aggression (i.e., instrumental and planned aggressive behavior) (Orobio de Castro, Merk, Koops, Veerman, & Bosch, 2005; but see Orobio de Castro et al., 2002). However, no training studies to date have differentiated between these two types of aggression. Second, studies in the anxiety domain have shown that training results in larger effects with increased numbers of training sessions (Menne-Lothmann et al., 2014). Therefore, compared to the study of Vassilopoulos et al. (2015), who presented 45 training trials spread over three sessions, we increased both the number of sessions and the num-

ber of trials per session. To counter the risk of the training becoming repetitive or monotonous, we also added novel audio–visual elements to the training.

Using five sessions of ambiguous scenario completion training, we trained adolescents selected for high levels of aggression to make more benign interpretations of others' intentions, and we compared the effects of this training on both hostile attribution bias and aggression with a test–retest control group. We expected that the training would result in a decrease in hostile attribution bias and that training-induced reductions in hostile attribution bias would primarily affect reactive aggression as opposed to proactive aggression.

## Method

### *Participants*

Participants were recruited from a Dutch secondary school for adolescents with average intelligence, specialized in the education of adolescents with either learning difficulties or social–emotional problems. A total of 65 adolescents were considered for inclusion based on teachers' informal assessments of their reactive aggression. Of these, 12 adolescents were excluded because they had a high likelihood of having autism spectrum disorder. For 47 of the remaining 53 adolescents, we obtained informed consent from both the adolescents and their parents. From this pool, we randomly invited 40 adolescents to participate. One participant dropped out after two training sessions and was not included in any of the analyses, resulting in a final sample of 39 adolescents ($M_{age}$ = 14.03 years, *SD* = 1.22, range = 12–16, 25 boys; training group: $n$ = 19, $M_{age}$ = 14.05 years, *SD* = 1.18, 14 boys; control group: $n$ = 20, $M_{age}$ = 14.00 years, *SD* = 1.30, 11 boys). Our sample's average IQ as measured with the Raven's Progressive Matrices test (Raven, Raven, & Court, 2003) was 99.85 (*SD* = 19.02).

### *Questionnaires*

To measure aggressive behaviors, we used the Dutch translation of the Reactive Proactive Questionnaire (RPQ; Cima, Raine, Meesters, & Popma, 2013). This questionnaire consists of 23 items divided over a reactive (e.g. "How often have you reacted angrily when provoked by others?") and a proactive (e.g. "How often have you taken things from other students?") aggression subscale. Due to an experimenter error, participants responded on 4-point Likert scales anchored with *almost never, sometimes, often,* and *almost always,* whereas the original RPQ (Cima et al., Raine et al., 2006) consists of 3-point Likert scales anchored with *almost never, sometimes,* and *often.* To make our scores comparable to studies that use the original scale, we transformed *almost always* responses on our scale to *often.*[1] Cronbach's alphas for the proactive and reactive subscales and the entire scale in our current study were .81, .86, and .89 pre-training and .61, .86, and .84 post-training, respectively.

### *Hostile attribution bias assessment*

To measure hostile attribution bias, we used a variant of the interpretation recognition task (Houtkamp, van der Molen, Salemink, de Voogd, & Klein, 2017). In this task, we presented 10 ambiguous scenarios of three sentences each, in which the motives or behaviors of others could be interpreted both positively and negatively. The scenarios were read out loud by the experimenter while participants were asked to imagine being in the situation. After each scenario, we portrayed the intentions or motives of the others in a negative/provocative versus positive/harmless manner, and we asked participants to rate the likelihood of them interpreting the motives in these ways on two separate 4-point Likert scales. Negative and positive interpretations were probed in a fixed randomized order. An example scenario reads as follows:

> You are in the changing room of your soccer club. Some of your teammates are talking outside. You hear someone mentioning your name.

[1] This transformation affected neither the significance nor the pattern of our results.

How big is the chance of you thinking, "My teammates are gossiping; they are probably saying mean things about me"?

How big is the chance of you thinking, "My teammates are enthusiastic and are probably discussing the tactics for the upcoming game"?

The same bias assessment task was used before and after the training. Cronbach's alphas for positive and negative interpretations were .77 and .77 pre-training and .82 and .72 post-training, respectively. We calculated hostile attribution bias scores as the difference between the average likelihood ratings for making positive/innocent interpretations and the average likelihood ratings for making negative/hostile interpretations. Positive attribution bias scores thus reflect a tendency to interpret ambiguity in a benign manner rather than a hostile manner, whereas negative scores reflect a tendency to interpret ambiguity in a hostile manner rather than a benign manner.

*Hostile attribution bias modification: Scenario completion training*

We presented ambiguous scenarios in which the motives or behaviors of others could be interpreted both positively and negatively, one sentence at a time, on a computer screen. The contents of the scenarios were inspired by the Novaco Anger Scale and Provocation Inventory (Novaco, 2003) and the scenarios used by Hawkins and Cougle (2013), but they were adapted to adolescents (i.e., shorter sentences, easier wording). In the last sentence, a crucial disambiguating word was presented as a word fragment, and participants were required to complete this fragment. The word fragment always disambiguated the scenario in a positive or harmless manner. An example of a scenario is as follows:

You're at the tennis court. A player hits the ball hard against your head.

It hurts a lot. The player is . . .

inexp-rienc-d.

After completing the word fragment, participants responded to a yes/no comprehension question to ensure that they processed the content of the scenario. The comprehension questions always focused on the positive outcome of the scenario ("Did the player accidentally hit the ball against your head?"). Finally, on responding, we provided feedback on the comprehension question ("Yes, the player is inexperienced" or "Wrong, the player is inexperienced") and the next scenario started.

The entire training consisted of five sessions, spread over two weeks, with the time between training sessions depending on regular teaching schedules. Each session lasted about 20 min and consisted of 50 unique scenarios. Within each session, each sequence of 10 scenarios was followed by a self-paced break. In each sequence of 10 scenarios, 7 scenarios were presented simultaneously in writing and aurally, 2 scenarios were also accompanied by a picture matching the content of the scenario, and 1 scenario was presented as a video fragment. These audio–visual cues were added to make the training less monotonous and to help adolescents with problems in verbal skills to complete the training.

*Procedure*

The entire procedure was approved by the ethical committee of the VU University Amsterdam. The study took place at the adolescents' school during school hours over the course of 4 weeks. In the first week, all participants completed the pre-training assessment of aggression and hostile attribution bias,[2] after which they were randomly assigned to either the training group or the control group. During the second and third weeks, participants in the training group completed the training procedure as described above, whereas participants in the control group received no intervention. Finally, in the fourth week, participants completed the post-training assessment of aggression and attribution bias. All tasks were performed individually, and all measurements (both training and assessments) were supervised by adult experimenters.

[2] During the pre-training assessment, participants also completed a number of cognitive tasks, including a Stroop color naming test, the forward digit span test, and the listening recall test. Detailed descriptions of these tasks are provided by van der Molen, Henry, and Van Luit (2014), but discussion and analysis of these tasks is beyond the scope of this article.

## Results

### *Group characteristics, correlations, and baseline effects*

Descriptive statistics and correlations between different constructs at baseline are presented in Table 1. Endorsement of positive interpretations was relatively consistently related to less aggression. Endorsement of negative interpretations was only related positively to reactive aggression. Partial correlations between reactive aggression and attribution bias indices (controlling for proactive aggression) were − .29, .35, and − .36 for positive interpretations, negative interpretations, and hostile interpretation bias, respectively (all *p*s between .074 and .025). Inversely, partial correlations between proactive aggression and these three bias indices (controlling for reactive aggression) were nonsignificant (all *p*s > .86) and ranged between − .029 and .024. Groups did not differ significantly on any of the baseline measures, all *t*s < 1.17, all *p*s > .25 (Table 2).

### *Effects of interpretation bias modification on hostile interpretations*

Assessing the effects of training on hostile attribution bias, we entered the mean endorsement ratings in a mixed-design analysis of variance (ANOVA), with Valence (positive vs. negative) and Time (pre-training vs. post-training) as within-participants factors and Training Group (training vs. control) as a between-participants factor. We found a significant interaction between Valence and Time, $F(1,$

**Table 1**
Group-level baseline descriptive statistics and correlations among different measures.

| | *M* | *SD* | 2. | 3. | 4. | 5. | 6. |
|---|---|---|---|---|---|---|---|
| 1. Positive interpretations | 2.31 | 0.53 | −.55* | .86* | −.36† | −.31 | −.41† |
| 2. Negative interpretations | 2.39 | 0.55 | | −.88* | .31 | .20 | .32† |
| 3. Hostile attribution bias | −0.08 | 0.97 | | | −.37† | −.26 | −.40* |
| 4. RPQ reactive | 8.90 | 5.26 | | | | .54* | .96* |
| 5. RPQ proactive | 2.79 | 3.43 | | | | | .72* |
| 6. RPQ total | 11.69 | 7.83 | | | | | |

*Note.* RPQ, Reactive Proactive Questionnaire. Correlations are based on Spearman coefficients.
† $p < .05$ (two-tailed).
* $p < .01$ (two-tailed).

**Table 2**
Descriptive statistics per group at pre-training and post-training and correlations among difference scores from pre-training to post-training.

| | Pre-training | | | | Post-training | | | | Pre–post change correlations | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Training group | | Control group | | Training group | | Control group | | | | | | |
| | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* | 2. | 3. | 4. | 5. | 6. |
| 1. Positive interpretations | 2.41 | 0.43 | 2.22 | 0.60 | 2.63 | 0.44 | 2.27 | 0.64 | −.48* | .87* | −.10 | −.28 | −.22 |
| 2. Negative interpretations | 2.45 | 0.46 | 2.34 | 0.63 | 2.11 | 0.31 | 2.30 | 0.59 | | −.82* | .11 | .09 | .15 |
| 3. Hostile attribution bias | −0.04 | 0.81 | −0.12 | 1.11 | 0.52 | 0.68 | −0.03 | 1.05 | | | −.11 | −.23 | −.23 |
| 4. RPQ reactive | 7.89 | 5.63 | 9.85 | 4.83 | 5.11 | 4.23 | 9.10 | 5.12 | | | | .34† | .86* |
| 5. RPQ proactive | 2.63 | 3.85 | 2.95 | 3.09 | 1.26 | 1.52 | 2.45 | 2.56 | | | | | .72* |
| 6. RPQ total | 10.53 | 8.36 | 12.80 | 7.33 | 6.37 | 5.19 | 11.55 | 6.35 | | | | | |

*Note.* RPQ, Reactive Proactive Questionnaire. Pre–post change correlations are based on Spearman coefficients.
† $p < .05$ (two-tailed).
* $p < .01$ (two-tailed).

37) = 8.15, $p$ = .007, $f$ = 0.47, a marginal main effect of Training Group, $F(1, 37)$ = 3.18, $p$ = .083, $f$ = 0.29, and a crucial three-way interaction, $F(1, 37)$ = 4.10, $p$ = .050, $f$ = 0.33, suggesting that the training affected interpretations over time. No other effects approached significance, all $F$s < 1.25, all $p$s > .27.

Following up on the three-way interaction, between-group comparisons of the interpretation bias scores (i.e., the difference between the endorsement of positive versus hostile interpretations) revealed a marginal difference post-training, $t(37)$ = 1.89, $p$ = .066, $d$ = 0.61 (Table 2). Within-group comparisons across the two measurements showed that the interpretation bias in the control group remained stable, $F$ < 1, whereas the participants in the training group had a significantly more benign interpretation bias after the training than before, $F(1, 18)$ = 9.38, $p$ = .007, $f$ = 0.72. Finally, comparing interpretation bias scores against zero (=no bias), participants in the control group had no preference to interpret ambiguity in either a benign or hostile manner, neither pre-training nor post-training, both $t$s < 1, both $p$s > .63. Participants in the training group also had no bias pre-training, but they did have a significant bias to interpret ambiguity in a benign manner post-training, $t(18)$ = 3.31, $p$ = .004, $d$ = 0.76. In sum, training tended to have the expected effect on interpretation bias, inducing a more benign interpretive style in the training group, yet these results were not always statistically robust.

### *Effects of interpretation bias modification on aggression*

To assess the effects of training on reactive and proactive aggression, we entered the RPQ subscale scores in two separate Time × Training Group mixed-measures ANOVAs. As predicted, for reactive aggression, the crucial interaction was significant, $F(1, 37)$ = 4.37, $p$ = .043, $f$ = 0.34, qualifying the main effects of Time, $F(1, 37)$ = 13.17, $p$ = .001, $f$ = 0.60, and Training Group, $F(1, 37)$ = 3.84, $p$ = .058, $f$ = 0.32. Follow-up between-group comparisons showed similar levels of reactive aggression pre-training, $t(37)$ = 1.17, $p$ = .251, but lower levels of reactive aggression post-training in the training group, $t(37)$ = 2.65, $p$ = .012, $d$ = 0.85 (Table 2). Within-group comparisons showed that the control group did not change significantly from pre-training to post-training, $F(1, 19)$ = 1.85, $p$ = .190, whereas the training group showed a significant decrease in reactive aggression from pre-training to post-training, $F(1, 18)$ = 11.72, $p$ = .003, $f$ = 0.81. For the proactive aggression scale, there was only a marginally significant main effect of Time, $F(1, 37)$ = 3.89, $p$ = .056, $f$ = 0.32, but no other effects were significant, both $F$s < 1.

Although the results of the ANOVA reported above indicate that training affected levels of reactive aggression, we also checked whether changes in interpretation bias were related to changes in aggression. None of the correlations between pre- versus post-training difference scores of hostile attribution bias and pre-training and post-training difference scores of the RPQ scales reached statistical significance, all $\rho$s between − .28 and .15 and all $p$s > .08 (Table 2), illustrating that training-induced changes in attribution bias were not significantly related to changes in aggression on an individual level.

## Discussion

Testing the effects of hostile attribution bias modification, we found that completing five relatively short training sessions induced a more benign attribution bias, although comparisons with the control group were only marginally significant. Crucially, compared with the control group, the training also resulted in decreased levels of self-reported reactive aggression but not proactive aggression, suggesting a causal relation between hostile attribution bias and aggression. Our results thus add empirical weight to theories attributing (reactive) aggression at least partly to hostile interpretations of others' intentions (Dodge, 2006). However, neither of our groups displayed a hostile attribution bias before the training. This may be due to the precise formulations or the content of the scenarios that we used to assess bias. Alternatively, in the absence of a nonaggressive control group, aggressive adolescents may differ from nonaggressive peers in that they are less likely to interpret others' intentions as benign rather than interpret others' intentions as hostile (but see Vassilopoulos et al., 2015).

Our results are largely compatible with the findings of Vassilopoulos et al. (2015), who used a similar scenario-based procedure, and the findings of Penton-Voak et al. (2013, Experiment 2), who used a facial emotion recognition training. We further extend these earlier findings by confirming that hostile attribution bias is especially related to reactive aggression (e.g., Orobio de Castro et al., 2005) and, thus, that hostile attribution bias retraining only results in changes in reactive aggression but not proactive aggression. However, even though our sample was slightly larger than the sample tested by Vassilopoulos et al. (2015) and we included more and longer training sessions, our medium-sized crucial interaction effects were smaller than the large-sized interaction effects reported by Vassilopoulos et al. It is possible that our attempts to make the training more accessible by incorporating auditory and visual cues unwittingly reduced its effectiveness (but see, e.g., Standage, Ashwin, & Fox, 2009). In addition, a correlational analysis showed that —on an individual level— changes in hostile attribution bias were not related to changes in aggression, thereby warranting a cautious interpretation of our findings. More and larger studies are warranted to provide reliable estimates of the effects of different types of hostile attribution bias retraining and to uncover the mechanisms underlying change in aggression.

Our sample was selected based on teacher reports of elevated aggression, but participants were not clinically referred. Results of studies with samples displaying more severe cases of aggression are mixed, with Penton-Voak et al. (2013) finding training-induced reductions in both hostile attribution bias and aggression, but Hiemstra et al. (2019) finding training-induced reductions only in hostile attribution bias but not aggression. Thus, although it seems relatively straightforward to change hostile attribution bias using cognitive training approaches, impacting on measures of aggression poses much more of a challenge, especially so in clinical populations. In this respect, it is worth noting that both Penton-Voak et al. (2013) and Hiemstra et al. (2019) used the facial emotion recognition training procedure. Some evidence from interpretation bias modification studies in the field of anxiety suggests that a good match between the content of the training and the outcome behavior is crucial for training-induced effects in bias to transfer to emotional outcomes (Mackintosh, Mathews, Eckien, & Hoppitt, 2013). Compared with training individuals to interpret ambiguous facial expressions as happy, scenario-based training may arguably provide more opportunities to carefully match the content of the training with the targeted change in aggression. Because scenario-based hostile attribution bias training has so far yielded relatively consistent results in nonclinical adolescent and student samples, it seems to be a suitable and promising candidate intervention for future studies in clinical or delinquent samples.

Our study also has limitations. One important issue concerns our no-training control group. Because the training was conducted during adolescents' school hours, it was ethically difficult to ask the control group to complete a placebo training that was not expected to have any effects. However, comparing the training with a simple test–retest control group means that we cannot unequivocally attribute our effects to the training alone. The effects may in part be due to being exposed to ambiguous scenarios or even to merely participating in a training protocol or being more exposed to experimenters. In addition, we assessed aggression only through self-report, which may have resulted in socially desirable answering. Furthermore, the internal consistency of the proactive subscale at post-training was poor. In future studies, aggression should be assessed using multiple measures and sources, including self-report, other-report, and objective observations of aggressive behaviors both in the lab and in real life to paint a more comprehensive picture of the effects of hostile attribution bias modification on aggression. Finally, we assessed only the short-term effects of the training. Thus, it remains unknown whether potential benefits of hostile attribution bias modification are long-lasting.

Notwithstanding these limitations, our study shows that hostile attribution bias can be changed through training and that such training may reduce self-reported reactive aggression. Thus, it seems worthwhile to further explore the applied potential of hostile attribution bias modification in aggressive adolescents and children.

## Acknowledgments

We thank Kirsten Hawkins and Jesse Cougle for sharing their stimulus materials and thank Lisanne Koppenberg and Helma Soeteman for helping with the data collection.

## References


Cima, M., Raine, A., Meesters, C., & Popma, A. (2013). Validation of the Dutch Reactive Proactive Questionnaire (RPQ): Differential correlates of reactive and proactive aggression from childhood to adulthood. *Aggressive Behavior, 39*, 99–113.

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

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.

Hiemstra, W., Orobio de Castro, B., & Thomaes, S. (2019). Reducing aggressive children's hostile attributions: A cognitive bias modification procedure. *Cognitive Therapy and Research, 43*, 387–398.

Houtkamp, E. S., van der Molen, M. J., Salemink, E., de Voogd, E. L., & Klein, A. M. (2017). Interpretation biases in socially anxious adolescents with a mild intellectual disability. *Research in Developmental Disabilities, 67*, 94–98.

Hudley, C., & Graham, S. (1993). An attributional intervention to reduce peer-directed aggression among African-American boys. *Child Development, 64*, 124–138.

Krebs, G., Pile, V., Grant, S., Degli Esposti, M., Montgomery, P., & Lau, J. Y. F. (2018). Research review: Cognitive bias modification of interpretations in youth and its effect on anxiety: A meta-analysis. *Journal of Child Psychology and Psychiatry, 59*, 831–844.

Mackintosh, B., Mathews, A., Eckien, D., & Hoppitt, L. (2013). Specificity effects in the modification of interpretation bias and stress reactivity. *Journal of Experimental Psychopathology, 4*, 133–147.

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

Menne-Lothmann, C., Viechtbauer, W., Höhn, P., Kasanova, Z., Haller, S. P., Drukker, M., ... Lau, J. Y. F. (2014). How to boost positive interpretations? A meta-analysis of the effectiveness of cognitive bias modification for interpretation. *PLoS One, 9* (6) e100925.

Nasby, W., Hayden, B., & Depaulo, B. M. (1980). Attributional bias among aggressive boys to interpret unambiguous social-stimuli as displays of hostility. *Journal of Abnormal Psychology, 89*, 459–468.

Novaco, R. W. (2003). *The Novaco Anger Scale and Provocation Inventory: Manual*. Los Angeles: Western Psychological Services.

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

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

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.

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

Raven, J., Raven, J. C., & Court, J. H. (2003). *Manual for Raven's Progressive Matrices and Vocabulary Scales*. San Antonio, TX: Harcourt Assessment.

Standage, H., Ashwin, C., & Fox, E. (2009). Comparing visual and auditory presentation for the modification of interpretation bias. *Journal of Behavior Therapy and Experimental Psychiatry, 40*, 558–570.

van der Molen, M. J., Henry, L., & Van Luit, J. E. H. (2014). Working memory development in children with mild to borderline intellectual disabilities. *Journal of Intellectual Disability Research, 58*, 637–650.

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