

Award Number: W81XWH-13-2-0001

**TITLE: : Development of Cognitive Bias Modification (CBM) Tools to Promote
Adjustment during Reintegration following Deployment**

PRINCIPAL INVESTIGATOR: Professor Yair Bar-Haim

CONTRACTING ORGANIZATION: Tel Aviv University, Address: Ramat-Aviv
Tel Aviv Israel, 69978

REPORT DATE: June 2017

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE June 2017		2. REPORT TYPE Final		3. DATES COVERED 15-OCT-2012 TO 14-OCT-2016	
4. TITLE AND SUBTITLE Development of Cognitive Bias Modification (CBM) Tools to Promote Adjustment during Reintegration following Deployment				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER W81XWH-13-2-0001	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Yair-Bar-Haim e-Mail: yair1@post.tau.ac.il.				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Tel Aviv University, Ramat-Aviv, Tel Aviv, Israel,69978				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT The overarching goal of the grant was to develop valid and reliable computerized tools to measure and modify anger-related cognitive biases and ultimately to examine their efficiency in reducing anger and adjustment difficulties among soldiers. The first aim of the research, addressed in the first reported study, was to measure the associations between anger measures and biases in anger-related attention and interpretation. This aim was addressed in the completion of a correlational study, and the publication of a scientific report describing its findings. The second aim was to explore the effect of cognitive interpretation training on anger related outcomes. This aim was addressed in a second study reported here, which found evidence for the efficacy of computerized interpretation training in reducing anger-related interpretations of ambiguous faces, mitigating self-reported state anger and reducing anger-related displaced retaliation behavior in an interpersonal game. No effect on self-reported trait anger was found. A scientific report describing these findings has been prepared and submitted for publication, and is currently under review. The third aim was to explore the effect of attention bias modification (ABM) training on anger related outcomes. This aim was addressed in the third study, which found no evidence for the efficacy of computerized attention training in reducing anger-related attention bias, or on self-reported anger measures and physiological reactivity related to an annoying manipulation. There was a training effect on anger-related displaced retaliation behavior in an interpersonal game, but since no training-related cognitive change was evidenced, it could not be concluded that this group difference is related to change in attention patterns. Conclusions and recommendations for future research following these findings are included in this summarizing report.					
15. SUBJECT TERMS anger, aggression, attention bias, interpretation bias, cognitive bias modification					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER (include area code)
U	U	U	UU	60	USAMRMC

Table of Contents

	<u>Page</u>
Introduction.....	4
Body.....	5
Key Research Accomplishments.....	6
Reportable Outcomes.....	7
Conclusion.....	11
References.....	12
Appendix A.....	13
Appendix B.....	27

INTRODUCTION

To enhance military performance in combat, soldiers learn to selectively attend to potential threats and to weigh any ambiguous information in the context of potential life-threatening danger. The development of such cognitive biases is expected to enhance soldiers' life preserving actions that among others include the use of combat-related aggressive action. Although the tendency to promptly and aggressively respond to potential threats in combat is crucial for survival, it may prove maladaptive in non-combat environments. Since deployed soldiers confront dramatic changes in environmental threat conditions, ranging from safety to acute danger, considerable plasticity in threat-related attention and threat interpretation is required. Insufficient plasticity in threat processing may confer risk for military performance and psychological adjustment both in theatre and upon reintegration back to civilian environments.

The overarching goal of the grant was to develop valid and reliable computerized tools to measure and modify anger-related cognitive biases and ultimately to examine their efficiency in reducing anger and adjustment difficulties among soldiers. This goal was perused through unique research collaboration between WRAIR and Tel Aviv University offering a combination of experts in advanced psychological research in military context and in translational cognitive-neuroscience research.

BODY

Recruitment of participants and data collection:

For details on recruitment of participants for the correlational study see the **Method** section in **Appendix A**.

For the interpretation training trial, a collaboration was formed with Bristol University in order to boost the number of participants and test alternative methods, thus the study involved two samples – our sample at Tel Aviv University and a second sample at Bristol University (See the **Method** section in **Appendix B** for details regarding this recruitment procedure).

In a third study – attention bias modification (ABM) training trial, data collection included 80 undergraduate students with high levels of self-reported trait anger.

Data analysis:

All data from the three studies has been analyzed, see reports in Appendices A and B.

Publications:

A scientific paper describing the findings from the first correlational study has been written and was recently published in *Cognition and Emotion* (Maoz, K., Adler, A.B., Bliese, P.D., Sipos, M.L., Quartana, P.J., Bar-Haim, Y., 2016. Attention and interpretation processes and trait anger experience, expression, and control. *Cognition and Emotion*, 1-12).

A scientific paper describing findings from the second study (cognitive interpretation training trial) has been written and is currently under review in *Emotion* (Maoz, K., Dalili, M.N., Adler, A.B., Sipos, M.L., Bliese, P.D., Quartana, P.J., Pine, D.S., Leibenluft, E., Penton-Voak, I.S., Munafò, M.R. and Bar-Haim, Y., *under review*. Increasing positive interpretation of ambiguous faces reduces displaced interpersonal retaliation.)

Problem areas

No real problems were encountered in conducting the three studies.

KEY RESEARCH ACCOMPLISHMENTS

- A Scientific report describing the findings of the first correlational study has been published in *Cognition and Emotion*. (Appendix A)
- A Scientific report describing the findings of the second study (cognitive interpretation training trial) has been written and submitted to *Emotion*, and is currently under review (Appendix B)
- A PhD dissertation based on the three studies described above is currently being finalized by Keren Maoz (PhD Candidate).

REPORTABLE OUTCOMES

The key findings from the correlational study suggest that attention bias toward angry faces in the dot probe task is associated with higher trait anger and anger expression and with lower anger control-in and anger control-out. Also, the propensity to quickly interpret ambiguous faces as angry (interpretation bias) was associated with greater anger expression and its subcomponent of anger expression-out and with lower anger control-out. Interactions between attention and interpretation biases did not contribute to the prediction of any anger component suggesting that attention and interpretation biases may function as distinct mechanisms. For a detailed description of the findings see the **Results** section in **Appendix A**.

The key findings from the interpretation training trial suggest that computerized cognitive interpretation training is effective in reducing anger-related interpretations of ambiguous faces (**Appendix B**: Figure 1), and that this effect generalized to novel faces which were not part of the training protocol (**Appendix B**: Figure 3). Training did not affect self-reported trait anger, but did mitigate self-reported state anger. Moreover, after receiving an unfair offer in an ultimatum game, participants in the active training group showed less displaced anger retaliation toward a neutral player, as manifested in significantly fairer offers compared to the placebo training group (**Appendix B**: Figure 4). The two groups did not differ in their offers to a bluntly unfair player (direct retaliation). For a detailed description of the findings and statistics see the **Results** section in **Appendix B**.

Data from the ABM training trial suggest that the computerized cognitive attention training task used in this study did not result in differences in attention bias between the active training and placebo groups. The repeated-measures ANOVA on attention bias pre- and post-training yielded no main effects of time or group nor an interaction effect, all $F_s < 1.3$, all $p_s > 0.27$. No significant difference in attention bias was found between the groups pre-training, $t_{(76)} = 0.43$, $p > 0.6$, or post-training, $t_{(76)} = 1.2$, $p > 0.2$. Thus, we concluded that this type of training may not be effective in changing attention bias in this specific population, or that the measurement tools employed are not sensitive enough to detect such changes. As can be expected based on the

failure to demonstrate a training effect on attention bias, the analysis examining the generalization of such effect to a set of new faces also failed to find a significant difference between the two groups, $t(76) = 0.51$, $p > 0.6$.

The results showed no evidence of a training condition-related change in self-reported trait anger, anger expression and anger mood scores from pre- to post-training:

Trait anger scores: The repeated-measures ANOVA on trait anger scores pre- and post-training yielded no main effects of time or group nor an interaction effect, all $F_s < 2$, all $p_s > 0.17$. No significant difference in Trait anger scores was found between the groups at baseline ($t(76) = 0.77$, $p > 0.44$) or at post-training ($t(76) = 1.29$, $p > 0.2$).

Anger Expression index: The repeated-measures ANOVA on anger expression scores pre- and post-training yielded no main effects of time or group nor an interaction effect, all $F_s < 2.8$, all $p_s > 0.1$. No significant difference in anger expression index was found between the groups at baseline ($t(76) = 0.88$, $p > 0.3$) or at post-training ($t(76) = 0.02$, $p > 0.9$).

Anger mood scale scores: The repeated measures ANOVA on anger mood scores pre- and post-training yielded a main effect of time, $F_{(1,75)} = 6.832$, $p < 0.05$, $\eta^2_p = 0.083$, suggesting participants overall reported more current anger in the final session (perhaps due to boredom from repeating the task) . No group effect nor a group by time interaction effect were found, $F_s < 2$, $p_s > 0.16$. No significant difference in anger mood scale scores was found between the groups at baseline ($t(75) = 1.05$, $p > 0.29$) or at post-training ($t(76) = 0.59$, $p > 0.56$).

Like in the interpretation training trial, the active training group in the ABM training trial showed reduced anger displacement in the Ultimatum Game. Specifically, after receiving an unfair offer in the game, participants in the active training group showed less displaced anger

retaliation toward a neutral player, as manifested in significantly fairer offers compared to the placebo training group, $t_{(76)} = 2.17, p < .05$. The two groups did not differ in their offers to the bluntly unfair player (direct retaliation), $t_{(76)} = 0.04, p > 0.9$, nor in their initial baseline offers, $t_{(76)} = 0.48, p > 0.63$ (see Figure 1). However, since here we failed to demonstrate a training-related cognitive change, it could not be concluded that this group difference in displaced anger is related to change in attention patterns.

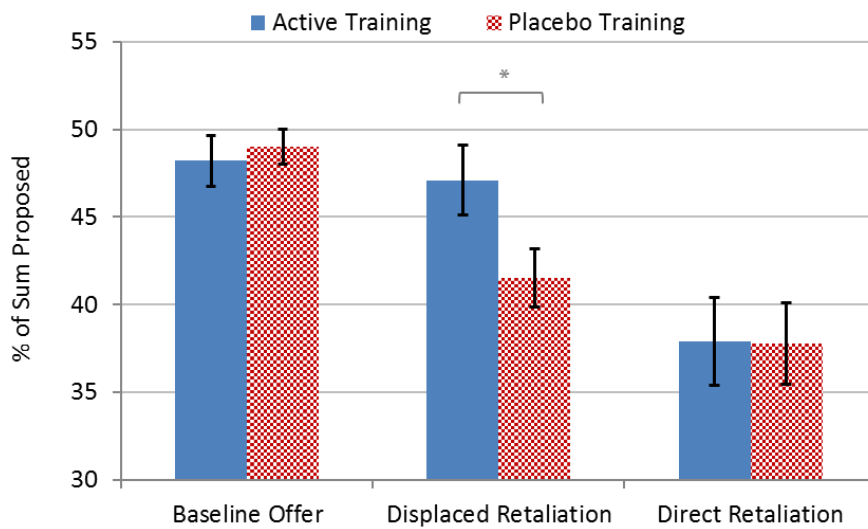


Figure 1: Mean percent of the sum proposed by participants in the active and placebo training groups in each round of the Ultimatum Game. Error bars represent standard errors.

Skin conductance level (SCL) in response to staged annoying manipulation

The repeated-measures ANOVA on SCLs before, during and after an annoying staged manipulation in the lab yielded a main effect of time, $F_{(2,140)} = 8.19, p < 0.001, \eta^2_p = 0.105$, and a marginally significant main effect of group, $F_{(1,70)} = 3.55, p = 0.064, \eta^2_p = 0.048$. No group by time interaction effect was found, $F_{(2,140)} < 1, p > 0.6$. The main effect of time indicates that the annoying manipulation indeed lead to significant elevations in mean SCL. To further explore this effect, 2 paired sample t-tests were conducted to compare SCL means before manipulation and

during manipulation, as well as before manipulation and after manipulation. These analyses demonstrated that mean SCL during the annoying manipulation, as well as after the annoying manipulation were significantly higher compared to mean SCL before the manipulation, $t_{(72)} = 3.56, p < 0.001$, and $t_{(75)} = 3.31, p < 0.001$, respectively. This serves as a manipulation check, suggesting that as expected participants demonstrated a physiological reaction to the annoying manipulation. However, as mentioned above, this effect did not interact with training condition.

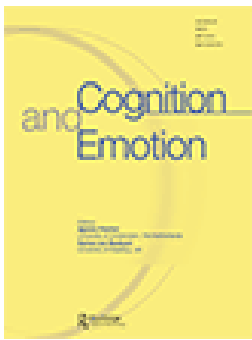
CONCLUSIONS

- Faster attention orientation and faster negative interpretation of threatening stimuli, as compared to non-threatening stimuli, are associated with greater self-reported anger, with differential patterns of associations between the two cognitive processes and the sub-components of anger experience, expression, and regulation. Interactions between attention and interpretation biases did not contribute to the prediction of any anger component suggesting that attention and interpretation biases may function as distinct mechanisms.
- It appears that interpretation biases and their modification protocols offer a more stable and reliable target for future studies and potential implementation for soldiers with anger expression and anger control difficulties.
- It seems that these cognitive modification protocols have stronger effect on behaviors related to displaced anger, compared to direct anger. Such an effect may be potentially important in the context of soldiers returning from deployment, who often deal with anger related to their combat experience. The computerized interpretation intervention may help preventing this anger from being maladaptively displaced and expressed within the family or community.
- If one well-powered RCT is to be funded to pursue the findings of the current project, it appears that:
 - a) Either the Bristol or TAU versions of the morph interpretation bias modification task is a viable selection. An advantage of the Bristol version is that it is shorter. An advantage of the TAU version is that it seems to yield a larger effect size for generalization, yet both tasks demonstrated significant generalization effects (see **Appendix B**: Figure 3).
 - b) The RCT could be applied in the context of anger management treatments for soldiers.
 - c) Outcomes of such RCT would ideally include indirect (behavioral, physiological) measures as well as assessments from independent evaluators in addition to self-reports of anger.
 - d) It is recommended to add a measurement session after about a month following the end of training to better examine trait-related changes and the stability of the cognitive and behavioral changes.

REFERENCES

Maoz, K., Adler, A.B., Bliese, P.D., Sipos, M.L., Quartana, P.J., Bar-Haim, Y., 2016. Attention and interpretation processes and trait anger experience, expression, and control. *Cognition and Emotion*, 1-12.

APENDIX A: a scientific report describing the findings of the correlational study. Published in *Cognition and Emotion*, 2016.



Attention and interpretation processes and trait anger experience, expression, and control

Keren Maoz, Amy B. Adler, Paul D. Bliese, Maurice L. Sipos, Phillip J. Quartana & Yair Bar-Haim

To cite this article: Keren Maoz, Amy B. Adler, Paul D. Bliese, Maurice L. Sipos, Phillip J. Quartana & Yair Bar-Haim (2016): Attention and interpretation processes and trait anger experience, expression, and control, *Cognition and Emotion*, DOI: [10.1080/02699931.2016.1231663](https://doi.org/10.1080/02699931.2016.1231663)

To link to this article: <http://dx.doi.org/10.1080/02699931.2016.1231663>

 [View supplementary material](#) 

 Published online: 22 Sep 2016.

 [Submit your article to this journal](#) 

 [View related articles](#) 

 [View Crossmark data](#) 

Attention and interpretation processes and trait anger experience, expression, and control

Keren Maoz^a, Amy B. Adler^b, Paul D. Bliese^b, Maurice L. Sipos^b, Phillip J. Quartana^b and Yair Bar-Haim^c

^aSchool of Psychological Sciences, Tel Aviv University, Tel Aviv, Israel; ^bCenter for Military Psychiatry and Neuroscience, Walter Reed Army Institute of Research, Silver Spring, MD, USA; ^cSchool of Psychological Sciences and Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, Israel

ABSTRACT

This study explored attention and interpretation biases in processing facial expressions as correlates of theoretically distinct self-reported anger experience, expression, and control. Non-selected undergraduate students ($N=101$) completed cognitive tasks measuring attention bias, interpretation bias, and Spielberger's State-Trait Anger Expression Inventory (STAXI-2). Attention bias toward angry faces was associated with higher trait anger and anger expression and with lower anger control-in and anger control-out. The propensity to quickly interpret ambiguous faces as angry was associated with greater anger expression and its subcomponent of anger expression-out and with lower anger control-out. Interactions between attention and interpretation biases did not contribute to the prediction of any anger component suggesting that attention and interpretation biases may function as distinct mechanisms. Theoretical and possible clinical implications are discussed.

ARTICLE HISTORY

Received 4 March 2016
Revised 18 August 2016
Accepted 27 August 2016

KEYWORDS

Anger; cognitive bias; attention; interpretation

Anger is a frequently experienced human emotion. However, the dispositional tendency to experience anger frequently and intensely is linked to a variety of deleterious outcomes. For example, high trait anger has been associated with elevated risk for cardiovascular problems (Chida & Steptoe, 2009; Smith, Glazer, Ruiz, & Gallo, 2004; Williams, 2010), as well as other physical illnesses (Suinn, 2001). Anger is correlated with health-risking behaviours, such as drinking, smoking, drug abuse, reckless driving, fighting, and unprotected sex (Adler, Britt, Castro, McGurk, & Bliese, 2011; Deffenbacher, Deffenbacher, Lynch, & Richards, 2003; Nichols, Mahadeo, Bryant, & Botvin, 2008; Sakusic et al., 2010), with risk for psychopathologies such as post-traumatic stress disorder (Feeny, Zoellner, & Foa, 2000; Meffert et al., 2008; Novaco, 2010), and with difficulties in spouse relationships (e.g. Baron et al., 2007), parent-to-child aggression (e.g. Mammen, Pilkonis, & Kolko, 2000), and workplace aggression (e.g. Hershcovis et al., 2007). However, the mechanisms underlying the tendency to experience

anger and how that anger is regulated are not well-understood.

Theoretical models suggest that cognitive processes play a pivotal role in the aetiology and maintenance of anger and aggression (e.g. Crick & Dodge, 1994; Wilkowski & Robinson, 2010). The Integrative Cognitive Model (ICM) of trait anger and reactive aggression (Wilkowski & Robinson, 2008; Wilkowski & Robinson, 2010), as well as the Social Information Processing (SIP) model (Crick & Dodge, 1994), postulate that both attention and interpretation processes reflect initial stages of social and affective information processing and that biases in these two processes are cognitive precursors for trait anger and reactive aggression. Amplified attention toward hostile and threatening cues and a predisposition to interpret ambiguous information in a hostile manner have both been linked to greater anger and aggression (Owen, 2011; Schultz, Grodack, & Izard, 2010). Anger-related stimuli induce greater attentional interference in participants with high versus low trait anger (e.g.

Van Honk, Tuiten, De Haan, Van den Hout, & Stam, 2001). And, anger has been associated with negatively biased interpretation of ambiguous information, including attributions of anger, hostility, and negative intent toward others (e.g. Schultz, Izard, & Bear, 2004; Wenzel & Lystad, 2005). However, to date, attentional and interpretational components of cognitive bias in anger have not been examined within-persons, hence their additive and multiplicative role in anger and anger regulation remain unknown.

Besides the lack of information about how these processes may function together, there are also some methodological gaps in the extant literature on cognitive biases in anger. Research on attention biases in the context of anger is scarce. Most studies have examined anger-related differences by measuring interference in emotional Stroop tasks (Eckhardt & Cohen, 1997; Van Honk, Tuiten, De Haan, et al., 2001; Van Honk et al., 2001), or attention biases to anger or aggression-related words in dot-probe tasks (Quartana, Yoon, & Burns, 2007; Smith & Waterman, 2003). Interestingly, within the field of anxiety a meta-analysis revealed that a significant emotional Stroop effect emerged using words as stimuli whereas a non-significant effect was noted when using pictures of faces. In contrast, in the dot-probe task significant anxiety-related effects emerged both for verbal and face stimuli (Bar-Haim, Lamy, Pergamin, Bakermans-Kranenburg, & van Ijzendoorn, 2007). Considering the extensive research on anxiety-related attention biases using facial stimuli in dot-probe tasks both in measurement (for a review see Bantini, Stevens, Gerlach, & Hermann, 2016) and in intervention (MacLeod & Mathews, 2012), it was surprising to find that only one study had used a faces-based dot-probe task in the context of anger. This study found attentional bias toward angry faces among youth with severe mood dysregulation (SMD), a condition characterised by high levels of irritability and anger (Hommer et al., 2014). In the current study, we used a version of the dot-probe task to measure selective attention to angry vs. neutral facial expressions as a function of anger.

Studies of interpretation bias have typically used verbal descriptions of social events to demonstrate that aggressive or angry individuals display a hostile attribution bias or a negative interpretation bias of ambiguous information (Orobio de Castro, Veerman, Koops, Bosch, & Monshouwer, 2002). However, fewer studies have used non-verbal stimuli such as facial expressions to examine anger-related interpretation biases (cf. Penton-Voak et al., 2013; Schultz et al.,

2004). In the current study we focus on potential interpretation biases in facial processing. One technique frequently used to examine interpretation biases of facial expressions relies on morphed faces (e.g. Pollak & Kistler, 2002; Richards et al., 2002). Morphing procedures typically mix two distinct facial expressions at either end of a continuum, thereby generating ambiguous facial stimuli along that continuum. It has been suggested that ambiguous faces containing conflicting information (e.g. morphing angry and happy expressions) may be effective at eliciting an interpretation bias because the images create a conflict in the cognitive classification of ambiguous expressions (Jusyte & Schönenberg, 2014). In the present study, we selected a task that required participants to label the emotion displayed by morphed ambiguous faces ranging from positive (happy) to negative (angry) emotion. This method has been used in the context of anxiety research (Garner, Baldwin, Bradley, & Mogg, 2009; Jusyte & Schönenberg, 2014; Maoz et al., 2016), but to our knowledge was applied only once in the context of anger (Penton-Voak et al., 2013).

Another limitation of the extant literature is related to the multifaceted nature of the construct of anger. That is, distinct cognitive biases might differentially relate to various components of how anger is expressed and experienced and these distinctions have rarely been clarified in research. A broad conceptualisation of anger and its management is illustrated in Spielberger's conceptualisation reflected in his State-Trait Anger Expression Inventory (STAXI-2; Spielberger, 1999). The STAXI-2 assesses experience, expression, and control of anger. State anger reflects a current emotional state marked by subjective feelings varying from mild irritation to intense rage. Trait anger is construed as a dispositional tendency to perceive situations as annoying or frustrating and respond with elevated state anger. Within this framework, state and trait anger are conceptually distinguished from anger expression. Anger expression is broadly divided into two types: anger expression-out, reflecting the tendency to outwardly express anger, and anger expression-in, reflecting the tendency to suppress anger or direct it towards oneself. Two additional sub-components of anger expression reflect an individual's ability to control and prevent anger expression toward the environment (anger control-out) and ability to control angry feelings by calming oneself down when angered (anger control-in) (Spielberger, 1999).

The current study builds on previous work by examining unique and common associations between components of anger and two putative cognitive processes, using measures of threat-related attention and interpretation biases. We assessed cognitive bias with two different tasks: (1) a dot-probe task that has been widely used to assess attention biases in anxiety disorders (Bar-Haim et al., 2007; MacLeod, Mathews, & Tata, 1986) but less frequently in the context of anger (e.g. Hommer et al., 2014); and (2) a task involving morphed faces designed to assess interpretation bias. We tested the associations between these bias assessments and self-reported trait anger, anger expression, and anger control. We predicted that both anger-related attention and interpretation biases would be associated with higher levels of trait anger and anger expression, and lower levels of anger control. We also hypothesised that anger will be most pronounced when individuals experience bias in both attention and in interpretation. It is not clear, however, whether attention and interpretation effects are additive or multiplicative; therefore, we also explored the contribution of potential interaction effects to anger beyond the separate effects of attention and interpretation biases alone.

Method

Participants

Undergraduate students from Tel Aviv University participated in the study ($N = 101$, mean age = 24.16 years, $SD = 2.22$; 65 females). The study was advertised across the campus, and students could enrol either on-line or via the phone. The only exclusion criterion was prior participation in other studies in our lab involving similar tasks. The study was approved by Tel Aviv University's Institutional Review Board. Participants provided signed informed consent and received either course credit or monetary compensation (\$5), to their preference.

Cognitive bias measures

Attention: the dot-probe task

The sequence of events on a dot-probe trial is described in Figure 15 (supplementary online material). Each trial began with the presentation of a fixation display (500 ms; white cross 1*1 cm), followed by a 500 ms presentation of an angry-neutral pair of

chromatic faces of the same actor. Each face appeared on a grey square background subtending 50 mm in width and 37.5 mm in height. The faces were presented with equal distance from the top and bottom of the fixation cross and separated by 15 mm. The top photograph was positioned 30 mm from the top edge of the screen. Faces were of 10 actors (5 females), taken from the NimStim stimulus set (Tottenham et al., 2009; models 1, 2, 3, 6, 7, 20, 24, 27, 31, 33). Following the faces display, a target probe (" $<$ " or " $>$ ", 4*4 mm) appeared at the location previously occupied by one of the faces. Participants were required to determine which direction the arrow pointed via pressing a keyboard button (" $<$ " or " $>$ ") using their dominant hand as quickly as possible, while avoiding errors. The target remained on the screen until response. The task was run using e-prime software (Psychology Software Tools, Pittsburgh, PA) and was displayed on a 15.6" monitor.

The task comprised 80 trials, displayed in a random order. Each of the 10 face pairs was presented 8 times. Across the eight repetitions of each face pair, position of the angry and neutral faces (above or below the fixation), target position (above or below the fixation) and probe type (" $<$ " or " $>$ ") were fully counterbalanced. Thus, the probe appeared equally in the location of angry and neutral faces. Anger-related attention bias was calculated by subtracting mean reaction time (RT) in trials where the target appeared at the angry face location from mean RT in trials where the target appeared at the neutral face location. Positive scores reflect attentional vigilance toward the angry faces, whereas negative scores reflect attentional threat avoidance (Bradley, Mogg, Falla, & Hamilton, 1998). The task took approximately five minutes to complete.

Interpretation: the emotion-perception task

Each trial began with a fixation display (800–1200 ms; white cross 1*1 cm), followed by a 200 ms chromatic presentation of a morphed face (90 mm in height and 70 mm in width). Face presentation time was selected to be similar to that used in previous studies in the context of anxiety (Maoz et al., 2016), and anger (Penton-Voak et al., 2013). Each face was followed by a 200 ms scrambled face mask. A question mark then appeared and remained on the screen until the participant determined whether the face was "angry" or "happy" by pressing one of two pre-specified keyboard buttons. Participants were instructed to use one hand for each button. The location of the

buttons (happy/angry) was counterbalanced across participants.

Morphed face sequences were generated from pictures of four actors (2 female), using Morpheus Photo Morpher v3.16. Each sequence consisted of 15 morphed faces ranging between the endpoints of happy and angry expressions of each actor. The endpoint faces were taken from the NimStim set (Tottenham et al., 2009; models 10, 18, 37, 41), and were different than the faces used in the dot-probe task. Each face was presented three times, for a total of 180 trials (4 sequences $\times$ 15 faces $\times$ 3 repetitions), displayed in random order. The task was run using e-prime software and was displayed on a 15.6" monitor. Figure 2S (supplementary online material) shows the sequence of events on a single trial and an example of one morphed sequence. The task took approximately eight minutes to complete.

Two measures were derived from this task, indexing two distinct aspects of interpretation bias: (a) percent of anger interpretations was used as an index of participants' tendency to interpret the face stimuli as angry; and (b) interpretation RT bias was calculated to index a propensity to make negative interpretations more quickly than positive interpretations. The interpretation RT bias was calculated by subtracting mean RT in angry interpretation trials from mean RT in happy interpretation trials (Maoz et al., 2016). Positive interpretation RT bias scores reflect a tendency to make angry interpretations faster than happy interpretations. Negative scores reflect the opposite tendency.

Self-reported anger

Self-reported anger was evaluated using the STAXI-2 (Spielberger, 1999). The STAXI-2 was translated to Hebrew in cooperation with the copyright owners (PAR Inc.). All participants in the study were fluent in Hebrew.

The state anger index, referring to current, situational anger, was not analysed for two reasons: (a) the current theoretical focus was on stable anger-related characteristics; and (b) because no manipulation of anger occurred, only negligible variability in state anger was expected (and observed).

The trait anger score was derived from a 10-item scale denoting the participant's disposition to perceive situations as annoying and react with anger elevations. The anger expression index combined responses to 32 items related to anger expression

and control. Specifically, this scale has four sub-components: Anger Expression-Out (AX-O) reflecting a tendency to express anger toward the environment; Anger Expression-In (AX-I) reflecting anger directed inward; Anger Control-Out (AC-O) reflecting the ability to control and prevent anger expression toward the environment; and Anger Control-In (AC-I) relating to the ability to control angry feelings by calming oneself down when angered. Higher anger expression index scores represent more anger expression (AX-I and AX-O) and less control over anger experience and expression (AC-I and AC-O). Items were rated on a Likert-type scale (1 = "almost never" to 4 = "almost always"). In the present sample, Cronbach's alphas for trait anger and for the anger expression index were 0.84 and 0.76, respectively. Cronbach's alphas for the anger expression sub-components were 0.77, 0.73, 0.93, and 0.92 for AX-O, AX-I, AC-O, and AC-I, respectively.

Procedure

Participants were given an explanation about the study and provided signed informed consent. The tasks were then performed in the following order: emotion-perception task, dot-probe task, and STAXI-2. A two-minute break was allowed between tasks.

Data analysis

Data cleaning

Dot-probe task. Trials with RT longer than 2000 ms, or incorrect response were excluded. Then, trials with RT deviating by more than three SDs from the mean of each participant were also excluded. This resulted in an average exclusion of 3.0% of all trials per participant.

Emotion-perception task. To gauge compliance with task demands, a threshold of 70% accuracy in the identification of the two overtly angry and overtly happy facial expressions was determined (Stoddard et al., 2016). Since each face was presented three times during the task, accuracy estimation was based on a total of 48 trials (4 extreme faces $\times$ 4 sets $\times$ 3 repetitions). One participant had an accuracy score lower than the 70% threshold and was removed from further analysis. Average accuracy score for the remaining 100 participants was 97.25% (SD = 3.5%).

Trials with RTs longer than 2000 ms were excluded from the interpretation RT bias calculation. Then, trials with RTs deviating by more than three SDs from the

mean of each participant were also excluded from the interpretation RT bias calculation. This resulted in the removal of an average of 3.5% of all trials per participant.

Regression analyses

To determine the contributions of attention and interpretation biases to the participants' trait anger and anger expression, we conducted two two-step hierarchical regression analyses. In the first step of each analysis, we entered attention bias, interpretation RT bias, and percent of anger interpretations as predictors. This allowed us to examine the relative contribution of the different cognitive biases to anger. In the second step, we added the two-way interactions (product terms) between the attention and interpretation biases. In the case of the anger expression index, which is composed of four theoretically distinguished sub-components, the significant overall model was followed-up by separate regressions for each of the sub-components. Analyses were conducted using IBM SPSS Statistics 22.

Results

Means, standard deviations, ranges, and correlations among the cognitive biases and self-reported anger variables are presented in Table 1. For scatter plots see Figure 3S in online supplementary materials. Mean trait anger and anger expression levels in the current sample were moderate (17.99 and 32.21, respectively) and similar to previously reported levels in normative adults samples (Spielberger, 1999). Since gender-based differences in some anger

components have been previously reported (Fischer & Evers, 2010), we tested for gender-based differences on all the measures used in our study. No gender-related differences were noted in any of the cognitive or self-reported anger variables ($ps > .10$).

Trait anger

The results of the linear regression model are summarised in Table 2. The first step significantly explained 9% of the variance in trait anger ($R^2 = 0.91$, $F_{(3,96)} = 3.20$, $p < .05$). However, only threat-related attention bias significantly predicted trait anger ($\beta = 0.26$, $p < .05$). The interactions between the different cognitive measures added in step two did not account for additional variance.

Anger expression index

The results of the linear regression model are summarised in Table 3. The first step significantly explained 13% of the variance in the anger expression index ($R^2 = 0.13$, $F_{(3,96)} = 4.65$, $p < .005$). Attention bias and interpretation RT bias each contributed to the prediction of anger expression index ($\beta = 0.24$, $p < .05$ and $\beta = 0.28$, $p < .01$, respectively) and uniquely contributed to the variability in anger expression with 5.6% and 7.1%, respectively. Percent of anger interpretations did not contribute to the variability in anger expression index scores. Interactions between the different cognitive measures added in step two did not account for additional variance.

Follow-up regression analyses examining the sub-components of the anger expression index revealed:

Table 1. Mean, SDs and range of the cognitive measures and self-report STAXI-2 scales, and Pearson correlations between the measures.

Measure	Mean	SD	Range	1	2	3	4	5	5a	5b	5c
1. Attention bias (dot-probe task)	6 ms	22	−36–74								
2. Interpretation RT bias (emotion-perception task)	−3 ms	46	−121–135	0.06							
3. Percent of anger interpretations (emotion-perception task)	54.6%	5.6	37.2–68.3	0.17 ⁺	0.35**						
4. Trait Anger (STAXI-2)	17.99	4.71	10–35	0.26**	0.16	0.08					
5. Anger Expression Index (STAXI-2)	32.21	13.45	5–69	0.23*	0.24*	−0.01	0.68**				
5a. Anger expression - Out	14.50	3.48	8–28	0.16	0.20*	−0.02	0.62**	0.78*			
5b. Anger expression - In	17.78	3.92	9–31	0.07	0.12	0.05	0.20*	0.39**	0.18		
5c. Anger control - Out	25.65	5.43	11–32	−0.19 ⁺	−0.24*	0.05	−0.65**	−0.84**	−0.68**	−0.01	
5d. Anger control - In	22.42	5.62	8–32	−0.22*	−0.14	0.01	−4.63**	−0.82**	−0.47**	−0.13	0.61**

⁺ $p < .08$.

* $p < .05$.

** $p < .01$.

Table 2. Estimated coefficients, standard errors, and 0.95 confidence intervals for predictors in the two steps of the regression model predicting trait anger.

	Predictor	Coefficients						Multicollinearity		Model	
		<i>B</i>	<i>SE</i>	<i>Beta</i>	<i>t</i>	<i>Sig</i>	95% CI	Tolerance	VIF	<i>R</i> ²	ΔR^2
Step 1	Attention bias	0.055	0.02	0.256	2.60**	<i>p</i> < .01	0.013–0.097	0.97	1.03	0.091*	–
	Interpretation RT bias	0.015	0.01	0.15	1.45	<i>n.s.</i>	–0.006–0.037	0.88	1.14		
	Percent of anger interpretations	–1.103	8.93	–0.01	–0.124	<i>n.s.</i>	–18.82–16.61	0.85	1.17		
Step 2	Attention bias	.064	.022	.295	2.836**	<i>p</i> < .01	0.019–0.108	0.88	1.13	0.110	<i>n.s.</i>
	Interpretation RT bias	.016	.011	.152	1.412	<i>n.s.</i>	–0.006–0.038	0.82	1.22		
	Percent of anger interpretations	–0.765	9.310	–0.009	–0.082	<i>n.s.</i>	–19.252–17.722	0.79	1.26		
	Attention bias × Interpretation RT bias	.001	.001	.126	1.144	<i>n.s.</i>	0.000–0.002	0.80	1.26		
	Attention bias × Percent of anger interpretations	–.021	.428	–.005	–.050	<i>n.s.</i>	–0.871–0.829	0.90	1.11		
	Interpretation RT bias × Percent of anger interpretations	–.241	.212	–.125	–1.139	<i>n.s.</i>	–0.662–0.179	0.79	1.26		

Note: *B* = unstandardised estimated coefficient; *SE* = standard error; CI = confidence interval.

**p* < .05.

***p* < .01.

Table 3. Estimated coefficients, standard errors, and 0.95 confidence intervals for predictors in the two steps of the regression model predicting anger expression index.

	Predictor	Coefficients						Multicollinearity		Model	
		<i>B</i>	<i>SE</i>	<i>Beta</i>	<i>t</i>	<i>Sig</i>	95% CI	Tolerance	VIF	<i>R</i> ²	ΔR^2
Step 1	Attention bias	0.148	0.060	0.240	2.478*	<i>p</i> < .05	0.029–0.266	0.97	1.03	0.127**	–
	Interpretation RT bias	0.083	0.030	0.284	2.786**	<i>p</i> < .01	0.024–0.143	0.88	1.14		
	Percent of anger interpretations	–37.51	24.97	–0.155	–1.503	<i>n.s.</i>	–87.07–12.04	0.85	1.17		
Step 2	Attention bias	.168	.062	.273	2.697**	<i>p</i> < .01	0.044–0.292	0.88	1.13	0.157	<i>n.s.</i>
	Interpretation RT bias	.089	.031	.304	2.896**	<i>p</i> < .01	0.028–0.151	0.82	1.22		
	Percent of anger interpretations	–29.321	25.857	–.121	–1.134	<i>n.s.</i>	–80.667–22.025	0.79	1.26		
	Attention bias × Interpretation RT bias	.001	.002	.048	.452	<i>n.s.</i>	–0.002–0.004	0.80	1.26		
	Attention bias × Percent of anger interpretations	.145	1.189	.012	.122	<i>n.s.</i>	–2.215–2.505	0.90	1.11		
	Interpretation RT bias × Percent of anger interpretations	–1.077	.588	–.196	–1.831	<i>n.s.</i>	–2.245–0.091	0.79	1.26		

Note: *B* = unstandardised estimated coefficient; *SE* = standard error; CI = confidence interval.

**p* < .05.

***p* < .01.

(a) anger suppression and anger expression toward oneself (AX-I) was not predicted by either attention or interpretation biases; (b) outward expression of anger (AX-O) was associated only with faster response to negative versus positive emotional interpretations (higher interpretation RT bias; $\beta = 0.24$, $p < .05$); (c) greater control over experienced angry feelings (AC-I) was predicted by less allocation of attention toward threat (lower attention bias; $\beta = -0.23$, $p < .05$); and (d) greater control over outward expression of anger (AC-O) was predicted by both lower attention bias toward threat ($\beta = -0.21$, $p < .05$) and lower interpretation RT bias ($\beta = -0.29$, $p < .01$).

Discussion

The current study explored the relations between cognitive processes (attention and interpretation biases) and self-reported anger measures reflecting theoretically distinguished aspects of anger experience, expression, and control. To our knowledge, this is the first study to examine the combined contributions of attention and interpretation biases on sub-components of self-reported anger expression and control. Results indicate a complex pattern of associations in which attention and interpretation biases each displayed unique associations with some anger components and overlapping contributions to other anger components. Unexpectedly, the interactions between the measured attention and interpretation biases did not contribute to the prediction of any anger component in the current sample.

Attention bias toward angry faces was associated with higher trait anger, whereas interpretation bias did not contribute to variability in trait anger. These findings suggest that the disposition to experience anger is related to a cognitive pattern of allocating more attention to threatening stimuli. This result converges with prior studies suggesting that individuals with high compared to low trait anger demonstrate greater attention bias toward negative stimuli (Van Honk, Tuiten, De Haan, et al., 2001). This finding is also consistent with the ICM framework (e.g. Wilkowski & Robinson, 2008, 2010), which suggests that high trait anger is associated with selective attention processes favouring hostile stimuli and specifically with difficulty in disengaging attention from such stimuli. These negatively biased attention patterns are thought to facilitate ruminative and hostile information processing that amplifies anger. In contrast, negative interpretation biases did not predict trait anger, a

result inconsistent with our hypothesis and with the ICM notion that hostile interpretations are of primary importance in anger elicitation.

With respect to the anger expression index, reflecting the balance between anger reactivity and anger regulation (Spielberger, 1999), the current findings revealed effects associated with both attentional and interpretational biases. Participants with high anger expression index displayed greater attention to negative faces on the dot-probe task and faster negative interpretations on the emotion-perception task. These results correspond with the ICM and SIP models, which posit that biases in early stages of attention, encoding, and interpretation of social information are precursors for reactive aggression.

Although attention and interpretation processes may both contribute to individual differences in anger expression, an inspection of the association between attention bias, interpretation RT bias, and the different sub-components of the anger expression index revealed a more specific pattern. While faster negative interpretation plays a role in the tendency to express anger toward the environment (AX-O), biased attention to angry faces is implicated in difficulties calming down when angered (AC-I). Both attention allocation toward threats and faster negative interpretations were associated with diminished ability to control one's outward anger expression (AC-O). Thus, while faster negative interpretation appears to be related to outward anger behaviour (i.e. aggression) and less control over such behaviour, biased threat-related attention is involved in regulation and control processes of both inner feelings and outward behaviour related to anger.

The finding that fast negative interpretation is associated with aggression-related anger components (AX-O, AC-O) is not surprising. Hostile attributional style has been identified as a precursor of aggressive behaviour development (Dodge, 2006), and aggressive individuals tend to demonstrate hostile attribution of intent (Orobio de Castro et al., 2002). Most of this literature focused on interpretation of social situations and many studies specifically focused on the attribution of negative intent to others. The current study focused on interpretation of anger in ambiguous facial expressions, demonstrating that fast-occurring interpretations of ambiguous faces as angry are related to self-reported anger expression toward the environment.

The finding that negative attention bias was associated with lower anger control (AC-I, AC-O)

corresponds with emotion regulation theories that include attention processes as key component of the emotion regulation system (Gross & Thompson, 2007). According to such models, distraction from negative thoughts serves as an early filter for blocking emotional information and preventing further emotion intensification (Sheppes & Gross, 2011). Similarly, the ICM model (Wilkowski & Robinson, 2008, 2010) suggests that self-distraction can reduce anger by reducing ruminative attention to hostile information. Biased attention to negative information may be a precursor for difficulties in employing distraction as an emotion regulation strategy. Therefore attention bias to negative stimuli may be related to lower control over anger reactivity, as indicated in the current study by the association between negatively biased attention and lower anger control (AC-I, AC-O).

Contrary to our expectations, no association was found between percent of anger interpretations and any of the self-reported anger measures. It may be that percent of anger interpretations, which is derived from a forced-choice response, is less sensitive than the RT-based measures. Additionally, percent of anger interpretations is based on explicit judgments and may reflect a conscious decision process. The two RT-based measures we used potentially reflect more implicit cognitive processes that may be less affected by conscious awareness. It also may be the case that this specific bias is more prominent in individuals with more extreme levels of anger, not represented in our non-selected sample, or individuals who have been exposed to extreme aggression, and thus may become sensitive to even mild expressions of anger, as has been found among abused children (Pollak & Kistler, 2002).

The current study did not find an effect for the interaction between attention and interpretation biases on anger measures. Typically, SIP and emotion regulation theories model attention and interpretation processes as inter-related rather than independent. Some see attention processes as preceding interpretation processes (Gross & Thompson, 2007; Sheppes & Gross, 2011), others see interpretation processes as influencing attention patterns (Wilkowski & Robinson, 2008, 2010), and some see these processes as having bidirectional influences (Crick & Dodge, 1994). Data from cognitive bias modification studies in anxiety support the notion that these two processes are influenced by one another. For example, participants trained to allocate attention

toward threat were later more likely to interpret ambiguous information in a threat-related manner (White, Suway, Pine, Bar-Haim, & Fox, 2011). Likewise, participants who completed interpretation training that encouraged benign, rather than negative, interpretations of ambiguous stimuli demonstrated greater ability to disengage attention from threat after training (Amir, Bomyea, & Beard, 2010). Our findings suggest that each of these processes has a distinct and independent contribution to anger (for similar findings in anxiety disorders see Pergamin-Hight, Bitton, Pine, Fox, & Bar-Haim, *in press*; Teachman, Smith-Janik, & Saporito, 2007) and have at least two implications. First, the findings suggest that individuals with biases in both attention and interpretation are not “more angry” beyond the effects of each component separately. That is, one might expect a “toxic” permutation in which (a) a propensity to quickly focus attention on anger-related stimuli combined with (b) a propensity to interpret ambiguous stimuli as anger-related leads to uncharacteristically high anger expression, but this did not appear to evident in the response patterns. The second implication associated with a lack of interactions is that one mechanism does not compensate or negate another. For instance, if an individual had the propensity to attend to anger-related stimuli, but was relatively unlikely to interpret ambiguous stimuli as anger-related, then one might expect to see an interaction suggesting suppressive effects. This type of pattern, however, was absent in the current study. We note that alternatively, this lack of interactive effect may be related to the moderate anger levels in the current sample. It may be the case that in participants with more extreme anger levels, not represented in the current sample, these two processes might interact and amplify the effect of each other.

The relative lack of anger-related research examining attention and interpretation biases in processing facial expressions opens up important and interesting areas for future research. First, we used only angry-happy morphed sequences in the emotion-perception task. By focusing on only one type of negative emotion, we were unable to ascertain whether anger was related to a specific tendency to interpret ambiguous faces as angry or, alternatively, a more general tendency to interpret ambiguous faces as negative. Future studies could explore this question by using stimuli consisting of various negative emotions (e.g. contempt, fear, and disgust), or ambiguous emotions (e.g. surprise), and tasks that include other negative

response options in addition to anger. Adding a neutral response option might also help differentiate whether the pattern of more negative interpretation was due to more anger interpretations, less happy interpretations, or both. Similarly, in the dot-probe task only angry-neutral face pairs were used and we were therefore unable to determine whether the bias was specifically related to angry faces, or alternatively related to a bias toward negative emotional faces or even to any emotional faces compared to neutral faces. To examine this, future studies may use stimuli sets consisting of various emotional expressions. Moreover, future research could elucidate whether anger is related to attention engagement by angry faces, difficulty disengaging attention from angry faces, or both (for similar examinations in anxiety see Grafton & MacLeod, 2014; Koster, Crombez, Verschuere, Van Damme, & Wiersema, 2006).

On a related note, we focused solely on anger thereby may have limited our ability to control for other traits that may be associated with the measured cognitive biases. For instance, similar attention and interpretation biases have been previously shown to be related to anxiety (e.g. Bradley et al., 1998; Maoz et al., 2016). Thus, an important and interesting direction for future research may be exploring concurrent and differential associations between cognitive biases, anger, and anxiety in the same sample. Some findings support associations between anxiety and anger (e.g. Kashdan & Collins, 2010) and it could be the case that similar cognitive biases in threat monitoring are related to both traits, reflecting two distinct emotional reactions to threat (anger-fight, fear-flight). Moreover, attention and interpretation biases in threat processing may be related to broader reactivity and regulation characteristics that are not only anger-related.

With respect to measurement-related future research, we note that although RT bias scores are commonly used in the cognitive bias literature, several studies have found the reliability of such subtraction-based scores in the dot-probe task to be low (Schmukle, 2005; Staugaard, 2009). Some alternative, more reliable measures for attention bias have recently been suggested, such as attention bias variability (Naim et al., 2015), ERP and fMRI signals (e.g. Kappenman, MacNamara, & Proudfit, 2015; White et al., 2016), and eye-tracking (e.g. Lazarov, Abend, & Bar-Haim, 2016; Shechner et al., 2013). Future studies may use these alternative measures to replicate the biases found in the current study.

Finally, future research should examine the generalizability of our findings. More specifically, anger levels in our sample were rather restricted and may have limited our ability to account for effects related to more extreme anger levels. Future studies could enrol individuals with elevated levels of anger, from samples of above-average self-reported anger, to samples of individuals who suffer from anger-related problems disrupting normative function (e.g. frequent anger outbursts, or repeated violent acts). Establishing causality between cognitive bias and anger, and establishing these cognitive biases among highly angry samples, could serve as preliminary steps for future cognitive bias modification protocols once stable targets for treatment emerge. Computerised cognitive modification protocols are increasingly studied in the context of treatment for anxiety disorders (for reviews see Bar-Haim, 2010; Beard, 2011; Hakamata et al., 2010; MacLeod & Mathews, 2012), yet the use of computerised cognitive modification protocols in the context of anger and aggression treatment is still scarce (cf. Hawkins & Cogle, 2013; Penton-Voak et al., 2013; Stoddard et al., 2016). Based on the findings of the current study, interventions focusing on attention and interpretation processes may potentially contribute to changes in different anger-related elements.

Targeting interpretation processes may help reduce expressions of anger. Indeed, interventions training toward benign rather than negative interpretation of social scenarios have been shown to reduce anger and aggressive behaviour (e.g. Hawkins & Cogle, 2013). To our knowledge, only one randomised controlled intervention study used computerised modification training to target interpretation of ambiguous faces (Penton-Voak et al., 2013). In this study, youth trained toward positive interpretation of ambiguous faces demonstrated less aggressive behaviour compared with a control group. Since the current study found anger expression to be associated with the more implicit aspect of negative interpretation RT bias rather than with the explicit percent of negative interpretations, cognitive modification effects on anger expression may potentially be enhanced by targeting interpretation response times in addition to interpretation judgments. Alternatively, targeting attention processes, and specifically reducing attention allocation to negative stimuli via attention bias modification may uniquely contribute to improvement in the capacity to control and relax anger feelings when experienced. To our knowledge,

no attention bias modification protocol has been used yet to target anger regulation. The current results suggest that both attention- and interpretation-based interventions could potentially contribute to better control over outward anger expression. This combined approach could be directly tested in future randomised controlled intervention studies.

The current study also had several limitations. First, the order of the two cognitive tasks was not counter-balanced across participants. This may have affected the findings by increasing emotional sensitivity during the morph task which might have had a carry-over effect during the dot-probe task. Future studies examining both cognitive biases should preferably use a design in which the two tasks are counterbalanced. Second, the current study is correlative, limiting inference regarding causality. The current analyses tested a theory-driven model suggesting that cognitive biases cause variations in anger levels. Nonetheless, it is also possible that individual differences in trait or state anger levels influence differential cognitive patterns. Although previous studies have demonstrated effect of changes in hostile interpretation on change in anger (e.g. Hawkins & Cogle, 2013), supporting the hypothesised causal sequence, this causal relation has not yet been demonstrated for attention biases. Future studies may test for causality by manipulating attention toward or away from hostile information while monitoring subsequent change in anger. A similar approach has been used to establish causality between attention bias and anxiety (Eldar, Ricon, & Bar-Haim, 2008; MacLeod, Rutherford, Campbell, Ebsworthy, & Holker, 2002). The alternative possible direction may be examined by manipulating state anger and measuring subsequent changes in cognitive biases. Third, the current study relied on self-reported anger, whereas future research may include more direct and objective measures of anger experience and expression (e.g. psychophysiological measures, criminal records).

In sum, this is the first study to explore attention and interpretation biases as well as their interaction as predictors of different self-reported anger sub-components. The current findings provide preliminary evidence for differential patterns of associations between cognitive processes and the sub-components of anger experience, expression, and regulation. These findings could be followed up using controlled experimental designs involving groups with heightened levels of trait anger or clinical groups, as well as cognitive modification protocols targeting different cognitive

mechanisms in order to further elucidate their specific causal contribution to anger and its regulation in anger-provoking conditions.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This study was supported by US Department of Defense [grant number W81XWH-13-2-0001].

References

- Adler, A. B., Britt, T. W., Castro, C. A., McGurk, D., & Bliese, P. D. (2011). Effect of transition home from combat on risk-taking and health-related behaviors. *Journal of Traumatic Stress*, 24(4), 381–389. doi:10.1002/jts.20665
- 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(2), 178–182. doi:10.1016/j.janxdis.2009.10.005
- Bantini, T., Stevens, S., Gerlach, A. L., & Hermann, C. (2016). What does the facial dot-probe task tell us about attentional processes in social anxiety? A systematic review. *Journal of Behavior Therapy and Experimental Psychiatry*, 50, 40–51.
- 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. doi:10.1111/j.1469-7610.2010.02251.x
- Bar-Haim, Y., Lamy, D., Pergamin, L., Bakermans-Kranenburg, M. J., & van Ijzendoorn, M. H. (2007). Threat-related attentional bias in anxious and nonanxious individuals: A meta-analytic study. *Psychological Bulletin*, 133, 1–24. doi:10.1037/0033-2909.133.1.1
- Baron, K. G., Smith, T. W., Butner, J., Nealey-Moore, J., Hawkins, M. W., & Uchino, B. N. (2007). Hostility, anger, and marital adjustment: Concurrent and prospective associations with psychosocial vulnerability. *Journal of Behavioral Medicine*, 30(1), 1–10. doi:10.1007/s10865-006-9086-z
- Beard, C. (2011). Cognitive bias modification for anxiety: Current evidence and future directions. *Expert Review of Neurotherapeutics*, 11(2), 299–311.
- Bradley, B. P., Mogg, K., Falla, S. J., & Hamilton, L. R. (1998). Attentional bias for threatening facial expressions in anxiety: Manipulation of stimulus duration. *Cognition & Emotion*, 12(6), 737–753.
- Chida, Y., & Steptoe, A. (2009). The association of anger and hostility with future coronary heart disease: a meta-analytic review of prospective evidence. *Journal of the American College of Cardiology*, 53(11), 936–946. doi:10.1016/j.jacc.2008.11.044
- Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social information-processing mechanisms in children's social adjustment. *Psychological Bulletin*, 115(1), 74–101. doi:10.1037/0033-2909.115.1.74
- Deffenbacher, J. L., Deffenbacher, D. M., Lynch, R. S., & Richards, T. L. (2003). Anger, aggression, and risky behavior: A comparison

- of high and low anger drivers. *Behaviour Research and Therapy*, 41(6), 701–718.
- Dodge, K. A. (2006). Translational science in action: Hostile attributional style and the development of aggressive behavior problems. *Development and Psychopathology*, 18(3), 791–814. doi:10.1017/s0954579406060391
- Eckhardt, C. I., & Cohen, D. J. (1997). Attention to anger-relevant and irrelevant stimuli following naturalistic insult. *Personality and Individual Differences*, 23(4), 619–629. doi:10.1016/s0191-8869(97)00074-3
- Eldar, S., Ricon, T., & Bar-Haim, Y. (2008). Plasticity in attention: Implications for stress response in children. *Behaviour Research and Therapy*, 46, 450–461. doi:10.1016/j.brat.2008.01.012
- Feeny, N. C., Zoellner, L. A., & Foa, E. B. (2000). Anger, dissociation, and posttraumatic stress disorder among female assault victims. *Journal of Traumatic Stress*, 13(1), 89–100. doi:10.1023/a:1007725015225
- Fischer, A. H., & Evers, C. (2010). Anger in the context of gender. In M. Potegal, G. Stemmler, & C. Spielberger (Eds.), *International handbook of anger* (pp. 349–360). New York, NY: Springer.
- Garner, M., Baldwin, D. S., Bradley, B. P., & Mogg, K. (2009). Impaired identification of fearful faces in generalised social phobia. *Journal of Affective Disorders*, 115(3), 460–465.
- Grafton, B., & MacLeod, C. (2014). Enhanced probing of attentional bias: The independence of anxiety-linked selectivity in attentional engagement with and disengagement from negative information. *Cognition and Emotion*, 28(7), 1287–1302.
- Gross, J. J., & Thompson, R. A. (2007). Emotion regulation: Conceptual foundations. In J. J. Gross (Eds.), *Handbook of emotion regulation* (pp. 3–24). New York, NY: Guilford.
- Hakamata, Y., Lissek, S., Bar-Haim, Y., Britton, J. C., Fox, N. A., Leibenluft, E., ... Pine, D. S. (2010). Attention bias modification treatment: A meta-analysis toward the establishment of novel treatment for anxiety. *Biological Psychiatry*, 68, 982–990. doi:10.1016/j.biopsych.2010.07.021
- Hawkins, K. A., & Cogle, J. R. (2013). Effects of interpretation training on hostile attribution bias and reactivity to interpersonal insult. *Behavior Therapy*, 44(3), 479–488.
- Hershcovis, M. S., Turner, N., Barling, J., Arnold, K. A., Dupré, K. E., Inness, M., ... Sivanathan, N. (2007). Predicting workplace aggression: A meta-analysis. *Journal of Applied Psychology*, 92(1), 228–238.
- Hommer, R. E., Meyer, A., Stoddard, J., Connolly, M. E., Mogg, K., Bradley, B. P., ... Brotman, M. A. (2014). Attention bias to threat faces in severe mood dysregulation. *Depression and Anxiety*, 31(7), 559–565.
- Jusyte, A., & Schönenberg, M. (2014). Threat processing in generalized social phobia: An investigation of interpretation biases in ambiguous facial affect. *Psychiatry Research*, 217(1), 100–106.
- Kappenman, E. S., MacNamara, A., & Proudfoot, G. H. (2015). Electrocortical evidence for rapid allocation of attention to threat in the dot-probe task. *Social Cognitive and Affective Neuroscience*, 10(4), 577–583.
- Kashdan, T. B., & Collins, R. L. (2010). Social anxiety and the experience of positive emotion and anger in everyday life: An ecological momentary assessment approach. *Anxiety, Stress, & Coping*, 23(3), 259–272.
- Koster, E. H., Crombez, G., Verschuere, B., Van Damme, S., & Wiersma, J. R. (2006). Components of attentional bias to threat in high trait anxiety: Facilitated engagement, impaired disengagement, and attentional avoidance. *Behaviour Research and Therapy*, 44(12), 1757–1771.
- Lazarov, A., Abend, R., & Bar-Haim, Y. (2016). Social anxiety is related to increased dwell time on socially threatening faces. *Journal of Affective Disorders*, 193, 282–288.
- MacLeod, C., & Mathews, A. (2012). Cognitive bias modification approaches to anxiety. *Annual Review of Clinical Psychology*, 8, 189–217.
- MacLeod, C., Mathews, A., & Tata, P. (1986). Attentional bias in emotional disorders. *Journal of Abnormal Psychology*, 95, 15–20. doi:10.1037//0021-843x.95.1.15
- 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. doi:10.1037//0021-843x.111.1.107
- Mammen, O. K., Pilkonis, P. A., & Kolko, D. J. (2000). Anger and parent-to-child aggression in mood and anxiety disorders. *Comprehensive Psychiatry*, 41(6), 461–468. doi:10.1053/comp.2000.16567
- Maoz, K., Eldar, S., Stoddard, J., Pine, D. S., Leibenluft, E., & Bar-Haim, Y. (2016). Angry-happy interpretations of ambiguous faces in social anxiety disorder. *Psychiatry Research*, 241, 122–127.
- Meffert, S. M., Metzler, T. J., Henn-Haase, C., McCaslin, S., Inslicht, S., Chemtob, C., ... Marmar, C. R. (2008). A prospective study of trait anger and PTSD symptoms in police. *Journal of Traumatic Stress*, 21(4), 410–416. doi:10.1002/jts.20350
- Naim, R., Abend, R., Wald, I., Eldar, S., Levi, O., Fruchter, E., ... Bar-Haim, Y. (2015). Threat-related attention bias variability and posttraumatic stress. *American Journal of Psychiatry*, 172(12), 1242–1250.
- Nichols, T. R., Mahadeo, M., Bryant, K., & Botvin, G. J. (2008). Examining anger as a predictor of drug use among multiethnic middle school students. *Journal of School Health*, 78(9), 480–486. doi:10.1111/j.1746-1561.2008.00333.x
- Novaco, R. W. (2010). Anger and psychopathology. In M. Potegal, G. Stemmler, & C. Spielberger (Eds.), *International handbook of anger* (pp. 465–498). New York, NY: Springer.
- Orobio de Castro, B., 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(3), 916–934.
- Owen, J. M. (2011). Transdiagnostic cognitive processes in high trait anger. *Clinical Psychology Review*, 31(2), 193–202. doi:10.1016/j.cpr.2010.10.003
- 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(5), 688–697.
- Pergamin-Hight, L., Bitton, S., Pine, D. S., Fox, N. A., & Bar-Haim, Y. (in press). Attention and Interpretation biases and attention control in youth with social anxiety disorder. *Journal of Experimental Psychopathology*. doi:10.5127/jep.053115
- Pollak, S. D., & Kistler, D. J. (2002). Early experience is associated with the development of categorical representations for facial expressions of emotion. *Proceedings of the National Academy of Sciences of the United States of America*, 99(13), 9072–9076. doi:10.1073/pnas.142165999

- Quartana, P. J., Yoon, K. L., & Burns, J. W. (2007). Anger suppression, ironic processes and pain. *Journal of Behavioral Medicine*, 30(6), 455–469.
- Richards, A., French, C. C., Calder, A. J., Webb, B., Fox, R., & Young, A. W. (2002). Anxiety-related bias in the classification of emotionally ambiguous facial expressions. *Emotion (Washington D C)*, 2(3), 273–287. doi:10.1037/1528-3542.2.3.273
- Sakusic, A., Avdibegovic, E., Zoricic, Z., Pavlovic, S., Gaspar, V., & Delic, A. (2010). Relationship between anger, alcoholism and symptoms of posttraumatic stress disorders in war veterans. *Medicinski Arhiv*, 64(6), 354–358.
- Schmukle, S. C. (2005). Unreliability of the dot probe task. *European Journal of Personality*, 19(7), 595–605.
- Schultz, D., Grodack, A., & Izard, C. E. (2010). State and trait anger, fear, and social information processing. In M. Potegal, G. Stemmler, & C. Spielberger (Eds.), *International handbook of anger: Constituent and concomitant biological, psychological and social processes* (pp. 311–325). New York, NY: Springer.
- Schultz, D., Izard, C. E., & Bear, G. (2004). Children's emotion processing: Relations to emotionality and aggression. *Development and Psychopathology*, 16(2), 371–388.
- Shechner, T., Jarcho, J. M., Britton, J. C., Leibenluft, E., Pine, D. S., & Nelson, E. E. (2013). Attention bias of anxious youth during extended exposure of emotional face Pairs: An eye-tracking study. *Depression and Anxiety*, 30(1), 14–21.
- Sheppes, G., & Gross, J. J. (2011). Is timing everything? Temporal considerations in emotion regulation. *Personality and Social Psychology Review*, 15(4), 319–331. doi:10.1177/1088868310395778.
- Smith, P., & Waterman, M. (2003). Processing bias for aggression words in forensic and nonforensic samples. *Cognition & Emotion*, 17(5), 681–701.
- Smith, T. W., Glazer, K., Ruiz, J. M., & Gallo, L. C. (2004). Hostility, anger, aggressiveness, and coronary heart disease: An interpersonal perspective on personality, emotion, and health. *Journal of Personality*, 72(6), 1217–1270. doi:10.1111/j.1467-6494.2004.00296.x
- Spielberger, C. D. (1999). *Professional manual for the state-trait anger expression inventory-2 (STAXI-2)*. Odessa, FL: Psychological Assessment Resources.
- Staugaard, S. R. (2009). Reliability of two versions of the dot-probe task using photographic faces. *Psychology Science Quarterly*, 51(3), 339–350.
- Stoddard, J., Sharif-Askary, B., Harkins, E. A., Frank, H. R., Brotman, M. A., Penton-Voak, I. S., ... Pine, D. S. (2016). An open pilot study of training hostile interpretation bias to treat disruptive mood dysregulation disorder. *Journal of Child and Adolescent Psychopharmacology*, 26(1), 49–57.
- Suinn, R. M. (2001). The terrible twos – anger and anxiety – hazardous to your health. *American Psychologist*, 56(1), 27–36. doi:10.1037//0003-066x.56.1.27
- Teachman, B. A., Smith-Janik, S. B., & Saporito, J. (2007). Information processing biases and panic disorder: Relationships among cognitive and symptom measures. *Behaviour Research and Therapy*, 45(8), 1791–1811.
- Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., ... Nelson, C. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research*, 168, 242–249. doi:10.1016/j.psychres.2008.05.006
- Van Honk, J., Tuiten, A., De Haan, E., Van den Hout, M., & Stam, H. (2001). Attentional biases for angry faces: Relationships to trait anger and anxiety. *Cognition & Emotion*, 15(3), 279–297. doi:10.1080/0269993004200222
- Van Honk, J., Tuiten, A., Van den Hout, M., Putman, P., De Haan, E., & Stam, H. (2001). Selective attention to unmasked and masked threatening words: relationships to trait anger and anxiety. *Personality and Individual Differences*, 30(4), 711–720.
- Wenzel, A., & Lystad, C. (2005). Interpretation biases in angry and anxious individuals. *Behaviour Research and Therapy*, 43(8), 1045–1054. doi:10.1016/j.brat.2004.02.009
- White, L. K., Britton, J. C., Sequeira, S., Ronkin, E. G., Chen, G., Bar-Haim, Y., ... Pine, D. S. (2016). Behavioral and neural stability of attention bias to threat in healthy adolescents. *NeuroImage*, 136, 84–93.
- 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(4), 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(1), 3–21. doi:10.1177/1088868307309874
- Wilkowski, B. M., & Robinson, M. D. (2010). The anatomy of anger: An integrative cognitive model of trait anger and reactive aggression. *Journal of Personality*, 78(1), 9–38. doi:10.1111/j.1467-6494.2009.00607.x
- Williams, J. E. (2010). Anger/hostility and cardiovascular disease. In M. Potegal, G. Stemmler, & C. Spielberger (Eds.), *International handbook of anger* (pp. 435–448). New York, NY: Springer.

APENDIX B: a scientific report describing the findings of the interpretation training trial.
Submitted to *Emotion*, 2017.

Emotion

Increasing positive interpretation of ambiguous faces reduces displaced interpersonal retaliation --Manuscript Draft--

Manuscript Number:	
Full Title:	Increasing positive interpretation of ambiguous faces reduces displaced interpersonal retaliation
Abstract:	This study explored the effect of two similar cognitive training tasks on interpretation of ambiguous emotional expressions, self-reported anger, and retaliatory behavior during interpersonal monetary decision making (Ultimatum Game). Participants were 160 Israeli and British undergraduates with high trait anger. Relative to placebo training, active training reduced anger-related interpretations of ambiguous faces, and this effect generalized to novel faces. Training did not affect self-reported trait anger, but did mitigate self-reported state anger. Finally, although participants in the active training condition did not differ from those in placebo training in direct retaliation against players who initiated unfair monetary transactions, the former showed less displaced retaliation, as reflected by fairer offers to a subsequent neutral player. These results suggest that in individuals with high trait anger, modifying interpretations of ambiguous emotional faces may reduce indirect behavioral manifestations of anger.
Article Type:	Article
Keywords:	displaced anger; interpretation; cognitive training; Ultimatum Game; faces
Corresponding Author:	keren maoz Tel Aviv University ISRAEL
Corresponding Author E-Mail:	maoz.keren@gmail.com
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Tel Aviv University
Other Authors:	Michael N. Dalili Amy B. Adler Maurice L. Sipos Paul D. Bliese Phillip J. Quartana Daniel S. Pine Ellen Leibenluft Ian S. Penton-Voak Marcus R. Munafò Yair Bar-Haim
Corresponding Author's Secondary Institution:	
First Author:	Keren Maoz
Order of Authors Secondary Information:	
Manuscript Region of Origin:	
Suggested Reviewers:	Emily Holmes Professor, Karolinska Institutet emily.holmes@ki.se Professor Holmes is a leading researcher in the field of cognitive biases and computer-

	based cognitive training protocols for various psychopathologies.
Opposed Reviewers:	
Order of Authors:	Keren Maoz
	Michael N. Dalili
	Amy B. Adler
	Maurice L. Sipos
	Paul D. Bliese
	Phillip J. Quartana
	Daniel S. Pine
	Ellen Leibenluft
	Ian S. Penton-Voak
	Marcus R. Munafò
	Yair Bar-Haim

Increasing positive interpretation of ambiguous faces reduces displaced interpersonal retaliation

Running Header: Interpretation training and displaced anger

Keren Maoz^{1,*}, Michael N. Dalili^{2,*}, Amy B. Adler³, Maurice L. Sipsos³, Paul D. Bliese^{3,4}, Phillip J. Quartana³, Daniel S. Pine⁵, Ellen Leibenluft⁵, Ian S. Penton-Voak^{2,**}, Marcus R. Munafò^{2,**} and Yair Bar-Haim^{1,6,**}

¹School of Psychological Sciences, Tel Aviv University

²School of Experimental Psychology, University of Bristol

³Center for Military Psychiatry and Neuroscience, Walter Reed Army Institute of Research

⁴Now at the Darla Moore School of Business, University of South Carolina.

⁵Emotion and Development Branch, National Institute of Mental Health

⁶Sagol School of Neuroscience, Tel Aviv University

* These authors made equal contributions to this study.

** These authors made equal contributions as principal investigators (PIs) of this study.

Corresponding Author: Correspondence concerning this article should be addressed to

Keren Maoz, School of Psychological Sciences, Tel Aviv University, Ramat Aviv, Tel Aviv 69987, Israel. E-mail: kerenmao@mail.tau.ac.il

Key words: displaced anger, interpretation, cognitive training, Ultimatum Game, faces

Abstract

This study explored the effect of two similar cognitive training tasks on interpretation of ambiguous emotional expressions, self-reported anger, and retaliatory behavior during interpersonal monetary decision making (Ultimatum Game). Participants were 160 Israeli and British undergraduates with high trait anger. Relative to placebo training, active training reduced anger-related interpretations of ambiguous faces, and this effect generalized to novel faces. Training did not affect self-reported trait anger, but did mitigate self-reported state anger. Finally, although participants in the active training condition did not differ from those in placebo training in direct retaliation against players who initiated unfair monetary transactions, the former showed less displaced retaliation, as reflected by fairer offers to a subsequent neutral player. These results suggest that in individuals with high trait anger, modifying interpretations of ambiguous emotional faces may reduce indirect behavioral manifestations of anger.

While anger is a universal human emotion (Plutchik, 2001; Scherer et al., 2004), failure to control anger can produce harmful aggressive or retaliatory behaviors. Some such behaviors may be expressed directly toward the anger-provoking agent, whereas other behaviors may be displaced and directed toward a different target, uninvolved in the original provocation. The idea of aggression displacement has roots in classic psychoanalytic theory (Freud, 1937; Sappenfield, 1954). It refers to altering the target of aggression as a way to avoid impulse expression toward "unsuitable" targets, which may result in unwanted consequences, or as a way to express the impulse when the original target is not available (Dollard et al., 1939; Marcus-Newhall et al., 2000). Yet displaced aggression may also result in harmful consequences. For example, employees who were "put down" by their supervisors tended to "put down" family members or domestic partners (Hoobler and Brass, 2006). Similarly, combat-related anger in soldiers has been associated with more aggressive tendencies, such as looking to start a fight after returning home from deployment (Adler et al., 2011).

Cognitive models of trait anger and aggression (e.g., Crick and Dodge, 1994; Wilkowski and Robinson, 2008, 2010; Wranik and Scherer, 2010) suggest that a predisposition to interpret ambiguous information and stimuli in a negative/hostile manner serves as one of the key cognitive precursors of high trait anger and reactive aggression. Empirical findings support the suggested association between anger and aggression and biased interpretation of ambiguous information (for reviews see Dodge, 2006; Orobio de Castro et al., 2002; Owen, 2011; Schultz et al., 2010). Following these lines of research, there have been recent attempts to moderate anger and aggressive behavior by modifying basic interpretive processes. For example, Penton-Voak et al. (2013) used a computerized feedback-based training to modify participants' interpretations of ambiguous facial expression from hostile-oriented (angry) to more positive-oriented (happy), resulting in reduced anger in

a sample of undergraduate students and in less aggressive behavior recorded by independent evaluators in a sample of frequently aggressive adolescents at risk for criminal behavior (Penton-Voak et al., 2013).

In the current study we wanted to explore the effect of computerized interpretation training on anger levels and on direct and displaced retaliation tendencies. One interesting paradigm from which to study both displaced and direct anger-related retaliation is within the context of the Ultimatum Game (Güth et al., 1982). In each round of the Ultimatum Game two players are asked to allocate a set amount of money. One player is assigned to be "the proposer", and has to offers how to distribute the money, while the other player is assigned to be "the responder" and has to choose whether to accept the offer (in which case the money is divided accordingly), or to reject the offer (in which case both players get nothing). With respect to direct anger, there is evidence that receiving an unfair offer in this game induces angry feelings that influence decision making and enhance direct retaliatory behavior, such as rejecting unfair offers or offering less money to anger- provoking counterparts (Fabiansson and Denson, 2012; Pillutla and Murnighan, 1996). Displaced anger has also been shown to influence decision making and retaliatory behavior in this paradigm. For example, participants who watched an anger-evoking video before playing the Ultimatum Game rejected more offers than participants who watched a happy mood video prior to the game (Andrade and Ariely, 2009). Presumably, watching an anger-evoking video primed participants to experience anger, which was subsequently displaced and resulted in more retaliatory behavior during the Ultimatum Game.

The aim of the current study was to examine the impact of two computerized cognitive interpretation training tasks on self-reported anger levels and on direct and displaced retaliation in the context of the Ultimatum Game. The sample consisted of students who had above-average trait anger (Israel site) or extreme trait anger (UK site). In a double

blind randomized controlled trial, we examined whether interpretation training designed to increase positive over negative perception of ambiguous facial expressions, compared to placebo training, resulted in modified interpretation patterns, reduced self-reported anger and less direct and indirect retaliation responses. We used two versions of interpretation training that were developed independently at the two universities, to assess whether one version would be more effective than the other in either engaging the cognitive target or in reducing anger and aggressive retaliation. We expected that both training versions would result in a change in interpretation pattern in the active training group but not in the placebo training group. Specifically, we hypothesized that following active training (in both task versions) participants would tend to interpret ambiguous faces as more positive compared to participants in the placebo training condition. We also expected that participants in the active training condition would demonstrate decreases in self-reported anger levels and less aggressive retaliation (both direct and indirect) compared to participants in the placebo training condition. We did not have an *a priori* hypothesis regarding which of the two training versions would be more efficacious.

Method

Design

The study used a 2 (task version: TAU/UoB) $\times$ 2 (training condition: Active/Placebo) $\times$ 2 (site: Israel/UK) between-subjects design.

Participants

The effect size for the change in interpretation bias in two previous studies ranged from $d=1.1-1.2$ (Penton-Voak et al., 2013, studies 1,2). However, the effect of training on self-reported anger among undergraduate students was smaller $d=0.84$ (Penton-Voak et al.,

2013, study 1). The power calculation suggested that 38 participants would be needed in each group to achieve 95% power to detect such an effect with an α level of 0.05. In order to account for a possible 5% dropout, and still have sufficient power to detect the expected effect on self-reported anger, we recruited 40 participants to each Training Condition $\times$ Task Version combination (20 of which were collected in each site).

Israel site: Eighty undergraduate students from Tel Aviv University participated in the study (20% males, mean age=23.48, SD=3.57). These participants were pre-selected based on their self-reported above-average trait anger levels on the State-Trait Anger Expression Inventory (STAXI-2; Spielberger, 1999), inclusion cutoff score ≥ 19 . A score of 19 reflects the 65th percentile of trait anger scores in normative adults (Spielberger, 1999).

UK site: Eighty undergraduate students from the University of Bristol participated in the study (33.75% males, mean age=21.05, SD=2.92). These participants were pre-selected based on their self-reported anger levels on the STAXI-2 completed on-line. Due to an error, instead of screening participants based on trait anger cutoff score ≥ 19 , the participants in Bristol were in fact screened based on a cutoff score ≥ 29 , corresponding to the 96th percentile of trait anger scores in normative adults (Spielberger, 1999).

The study was approved by both the Tel Aviv University Institutional Review Board and the University of Bristol Faculty of Science Research Ethics Committee. Participants provided signed informed consent and received either course credit or monetary compensation (80NIS or £20). In addition, participants received the amount of money they earned during the Ultimatum Game.

Interpretation training tasks

Stimuli

Stimuli for both task versions (TAU/UoB) were images of morphed faces, ranging from unambiguously happy to unambiguously angry images, with emotionally ambiguous images in the middle of the range. Both tasks were run using E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA).

TAU version: Morphed face sequences were generated using Morpheus Photo Morpher (version 3.16) from pictures of four actors (2 female), taken from the NimStim set (Tottenham et al., 2009; models 10, 18, 37, 41). Each morphed sequence consisted of 15 images ranging between the happy and angry expressions of each actor. Each trial began with a fixation cross (800-1200ms), followed by a 200 ms presentation of a morphed face image, a 200ms visual mask and a question mark that remained on the screen until response. In this version the measurement and training phases each consisted of 180 trials (4 actors $\times$ 15 morphed images $\times$ 3 repetitions), displayed in random order.

UoB version: Stimuli were generated in two steps, as described in Penton-Voak et al. (2013). First, established techniques (Tiddeman et al., 2001) were used to generate prototypical happy and angry composite images from 20 individual male faces showing a happy facial expression and the same 20 individuals showing an angry expression. The original images came from the Karolinska Directed Emotional Faces (Lundqvist et al., 1998). Then, the two "happy" and "angry" prototypical images were used as endpoints to generate a linear morph sequence consisting of 15 equally spaced images that were used as experimental stimuli. Each trial began with a fixation cross (1,500 to 2,500 ms), followed by a 150 ms presentation of a face image, a visual mask (150 ms) and a prompt asking the participant to respond. The baseline and post-training measurements were identical and consisted of 45 trials, with each of the 15 morphed stimuli presented three times in random order. The training phase consisted of six training blocks, with 30 trials each (two presentations of each morphed image, in random order), for a total of 180 training trials.

Training sequence

For both task versions each training session included three parts: baseline measurement, training, and post-training measurement.

Baseline measurement: Participants were told that on each trial a face would appear briefly, and were asked to determine as quickly as possible whether each face was "angry" or "happy" by pressing one of two pre-specified keyboard buttons. Two bias measures were calculated: a) percent of anger responses out of all the responses was calculated and transformed to an angry-happy balance point, reflecting an estimate of the point in the morphed sequence in which a participant's interpretation shifted from happy to angry interpretation (for a detailed description see Penton-Voak et al., 2013), and b) reaction time (RT) bias, which was calculated by subtracting mean RT in trials with "angry" responses from mean RT in trials with "happy" responses (Maoz et al., 2016a; Maoz et al., 2016b), in order to index a propensity to make negative interpretations more quickly than positive interpretations. RT bias scores above zero reflect a tendency to make angry interpretations faster than happy interpretations, whereas scores below zero reflect the opposite tendency.

Training phase: Each trial in the training phase was similar to a trial in the baseline measurement with respect to stimulus presentation characteristics and task demands. In the training phase, however, feedback was provided following the participant's response by presenting a message on the screen saying, "Correct/Incorrect! That face was happy/angry". The feedback was given based on the participant's condition (placebo/active training), which was randomly assigned at the beginning of the session. In the active training condition, feedback was based on the participant's baseline balance point, but the "correct" classification was shifted two morphed images toward the angry end of the continuum, so that the two ambiguous images nearest to the balance point that the participant had classified

as angry at baseline received feedback suggesting that happy was the correct response during the training phase. In the placebo condition, feedback was directly based on the participant's baseline balance point. Responses were classified as "correct" when participants identified face images between their original balance-point and the happy endpoint as happy, and faces between the balance point and the angry endpoint as angry. Therefore, no change in interpretation bias was expected in this condition.

Test measurement: The test measurement was identical to the baseline measurement. Following the test measurement, participants' percent of anger responses and RT bias scores were again calculated, to examine whether the training modified each of the bias indices.

Near-transfer of training

Before the training sequence in session 1, and after the training sequence in session 2, participants completed a generalization near-transfer measurement. This measurement was similar to the measurement phase in the training sequence, but consisted of morphed images that were not included in the training session. For this purpose, another morphed sequence was generated. This sequence also consisted of 15 morphed faces ranging between the endpoints of happy and angry expressions of a male actor from the NimStim set (Tottenham et al., 2009; model 34). This measurement consisted of one block with 45 trials, in which each of the 15 morphed stimuli was presented three times in a random order. All other characteristics of stimuli presentation were the same as those in the measurement phase of the relevant task version (TAU/UoB).

Self-reported anger

Self-reported trait anger was evaluated using the STAXI-2 (Spielberger, 1999). In UK, the English version was used; in Israel, a Hebrew translation was used. The STAXI-2

was translated to Hebrew in cooperation with the copyright owners and feedback from the original author (PAR Inc.). In the present sample, Cronbach's alphas in the Israel site were 0.74 and 0.75 for trait anger in sessions 1 and 2, respectively, and 0.66 and 0.77 for anger expression in sessions 1 and 2, respectively. Cronbach's alphas in the UK site were 0.67 and 0.74 for trait anger in sessions 1 and 2, respectively, and 0.67 and 0.76 for anger expression in sessions 1 and 2, respectively.

Self-reported online state anger was evaluated using an analog mood scale, on which participants were required to rate their current anger level on a scale ranging from 0 to 30 (Abend et al., 2014).

Ultimatum Game

A virtual Ultimatum Game was introduced to the participants as an interactive web application. The participants were told they would play an interactive web game with a number of other players (who were in fact virtual players). At the beginning of the game, pictures of the four "players" were presented on the screen. This included the participants' own picture as well as pictures representing each of the three other players. In each round of the game the participant was "randomly paired" with a single other player, and this player's photo was presented on the screen together with the participants' photo (as is commonly the setting in on-line chats or games). Participants were told that the responses of the responders would be revealed only at the end of the game, as opposed to after each round, and that the sums of money would be divided accordingly.

In each round, the pair of players had to share a predetermined sum of money (10 NIS in Israel and £5 in UK). The game consisted of four rounds. In the first round, the participant was assigned the role of the proposer and made an offer to the first unfamiliar player. We termed this round the "baseline proposal". In the second round the participant was assigned

the role of the responder and received an unfair offer (20% of the sum) from a new unfamiliar player, who we termed the "unfair" player. The participant then decided whether to accept or reject the unfair offer. In the third round, the participant was paired with the third unfamiliar player and was assigned the role of the proposer. We were interested in whether the participant's proposal to this "innocent" unfamiliar player would be affected by displaced retaliation following the unfair offer from the "unfair" player in the previous round. Our hypothesis was that participants in the active training condition will show less displaced retaliation toward the "innocent" player in this round reflected in more fair offers relative to the placebo condition. Finally, the fourth round again paired the participant with the "unfair" player, but the participant was assigned the role of "proposer", offering an opportunity for direct retaliation. Our hypothesis was that participants in the active training condition will show less direct retaliation, i.e. that their offers in this round will be more fair than the offers made by participants in the placebo condition.

Procedure

Participants arrived at the laboratory for two experimental sessions, one week apart. In the first session, participants were given an explanation about the study and provided signed informed consent. They next completed the two trait scales from the STAXI-2 questionnaire and rated their current anger level on an analog mood scale ranging from 0 to 30. Participants were then told they were going to perform an emotion recognition task, in which a face would be presented briefly in each trial. Participants were told to judge whether the face was angry or happy. The first generalization block was then run. After participants completed this block, they were told that they would perform the same task but with different faces. They were told that the task consisted of three parts, during one of which they would receive feedback regarding the accuracy of their responses. Then, the three-stage emotion perception training procedure began (baseline; training; test), with a short break separating

each block. After the task, participants again rated their current anger level on the computerized mood scale. The first session lasted 35-45 minutes.

Participants were told that the second session involved two parts: one that continued the prior week's session with computer tasks and questionnaires, and a second part, The Ultimatum Game, which was presented as a pilot study assessing decision making. During the first part, participants rated their current anger level on the computerized mood scale, completed the three-part interpretation training and an additional generalization block, again rated their current anger level on a mood scale and completed the STAXI-2 questionnaire. Subjects then completed The Ultimatum Game. After the Ultimatum Game participants completed a feedback form and received the money they earned in the game as well as the course credit or monetary compensation for participating in the study. The second session lasted approximately 55-60 minutes.

Data cleaning

Three participants (1 male) dropped out from the study before the second session and were thus removed from analyses. To eliminate subjects with poor compliance, a 70% accuracy threshold was set for identifying the two overtly angry and overtly happy facial expressions (Stoddard et al., 2016), which one participant failed to pass and was thus excluded from further analysis. The average accuracy score for the remaining 156 participants was 96.03% (SD=3.74%). Trials with RTs longer than 2000 ms were excluded from the RT bias calculation. Then, for each participant, trials with RTs deviating by more than 3 standard deviations from the mean of each response type (happy/angry) were also excluded. This resulted in the removal of 4.8% of all trials.

Data Analysis

Cognitive measures and trait anger: The two cognitive indices (percent of anger responses and RT bias) were submitted to $2 \times 2 \times 2 \times 2$ repeated-measures ANOVAs with Time (baseline, post-training) as a repeated within-subjects factor and Training Condition (active training, placebo training), Task Version (TAU, UoB) and Site (Israel, UK) as between-subjects factors. The same analysis was applied for the cognitive indices derived from the generalization blocks, as well as for the STAXI-2 measures of trait anger and anger expression.

Baseline proposals in the Ultimatum Game: To examine whether the training had any effect on baseline proposals in the Ultimatum Game, an ANOVA with Training Condition, Task Version, and Site as between-subjects factors was conducted, with percent of the sum proposed to the other player at baseline (round 1) as the dependent variable.

Direct retaliation in the Ultimatum Game: In the version of the ultimatum game we used, direct retaliation toward the "unfair" player could be expressed in two ways: first, by choosing to reject the unfair offer in round 2; and, second, by proposing an unfair offer when reencountering the "unfair" player in round 4. To examine whether training had an effect on direct retaliation behavior in the ultimatum game, we first used a Chi-squared test to examine whether the two training conditions differed in the proportion of participants who rejected the unfair offer. Second, we conducted an ANOVA with Training Condition, Task Version, and Site as between-subjects factors, and percent of the sum proposed to the "unfair" player (round 4) as the dependent variable.

Displaced retaliation in the Ultimatum Game: To examine whether training had an effect on displaced retaliation toward the "innocent" player in the ultimatum game, we conducted an ANOVA with Training Condition, Task Version and Site as between-subjects factors, and percent of the sum proposed to the scapegoat (round 3) as the dependent variable.

Change in anger mood within sessions: An anger mood change score was calculated within each training session by subtracting scores on the computerized anger mood scale before training from anger mood scores after training. The change scores were submitted to a $2 \times 2 \times 2 \times 2$ repeated-measures ANOVA with Session (first, second) as a repeated within-subject factor and Training Condition (active, placebo), Task Version (TAU, UoB) and Site (Israel, UK) as between-subjects factors.

Results

Means and 95% confidence intervals (CIs) of baseline and post-training measurements by Group are presented in Table 1.

Table 1. Means and 95% CIs of Baseline and Post-Training Measurements by Group.

	Active training group (N=77)		Placebo training group (N=79)	
	Baseline	Post-training	Baseline	Post-training
Gender (% Male)	37.5%		36.2%	
Age	21.83 [21.10, 22.56]		22.55 [21.75, 23.35]	
Cognitive task				
Anger responses (%)	54.95 [52.94, 56.96]	40.29 [37.83, 42.75]	55.86 [53.84, 57.88]	56.43 [54.03, 58.83]
RT bias (ms)	7 [-11, 26]	-58 [-79, -37]	-1 [-24, 22]	21 [-4, 36]
Generalization blocks				
Anger Responses (%)	55.38 [53.68, 57.08]	45.95 [43.64, 48.26]	55.61 [54.08, 57.14]	57.33 [55.18, 59.48]
RT bias (ms)	7 [-15, 29]	-29 [-53, -5]	21 [-6, 48]	41 [18, 63]
STAXI-2				
Trait Anger Score	25 [24, 26]	25 [24, 26]	25 [24, 26]	25 [24, 26]
Anger Expression Index	47 [44, 50]	47 [44, 50]	47 [45, 50]	47 [44, 50]

Cognitive change in interpretation bias

Percent of anger responses: The ANOVA indicated a main effect of Time, $F(1, 148) = 112.87$, $p < .001$, $\eta^2_p = .43$, and Training Condition, $F(1, 148) = 35.44$, $p < .001$, $\eta^2_p = .19$, which were subsumed under a Time $\times$ Training Condition interaction, $F(1, 148) = 129.72$, $p < .001$, $\eta^2_p = .47$ (Figure 1). Participants in the active training group interpreted fewer faces as

angry post-training relative to pre-training, $t(76) = 14.57, p < .001$, Cohen's $d = 1.70$; no such change was evident in the placebo group, $t(78) = 0.57, p > .250$ (for means and CIs see Table 1). There was no evidence that this interaction was moderated by Task Version or Site (Time $\times$ Training Condition $\times$ Site, and Time $\times$ Training Condition $\times$ Task Version interactions, $F_s < 1, p_s > .250$).

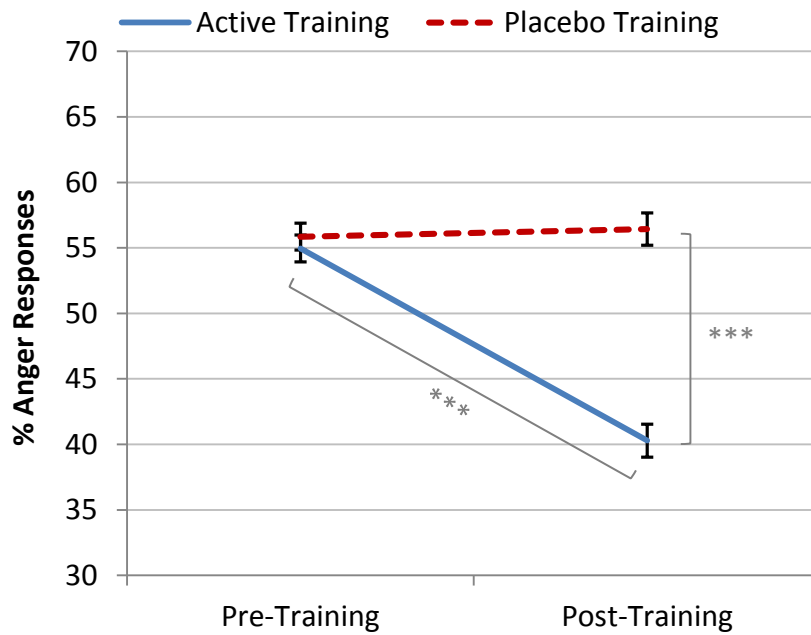


Figure 1: Mean percent of anger responses at pre- and post-training for the active and placebo training groups. Error bars represent standard errors.

RT bias: The ANOVA yielded main effects of Time, $F(1, 148) = 8.05, p = .005, \eta^2_p = .05$ and Training Condition, $F(1, 148) = 7.44, p = .007, \eta^2_p = .05$, which again were subsumed by a Time $\times$ Training Condition interaction, $F(1, 148) = 23.00, p < .001, \eta^2_p = .14$. The RT bias of participants in the active training group decreased from pre- to post-training, $t(76) = 5.61, p < .001$, Cohen's $d = .64$, but no such change occurred in the placebo training group, $t(78) = 1.34, p = .186$ (for means and CIs see Table 1). There was a Time $\times$ Training Condition $\times$ Site interaction effect, $F(1, 148) = 3.92, p = .050, \eta^2_p = .03$), suggesting

different patterns in Israel and UK sites. As can be seen in Figure 2, both samples manifested the training effect with a large effect size in Israel, $t(39) = 6.12, p < .001$, Cohen's $d = .97$ and a medium effect size in UK $t(36) = 2.61, p = .013$, Cohen's $d = .43$. No such reduction in RT bias scores was evident in the placebo training groups in both sites ($ps > .250$). Moreover, in both sites there was evidence that the two training conditions (active vs. placebo) differed in RT bias scores after training, $t(78) = 5.58, p < .001$, Cohen's $d = 1.25$ and $t(74) = 2.44, p = .017$, Cohen's $d = .56$ for the Israel and UK sites, respectively, but not before training ($ts < 1.42, ps > .16$).

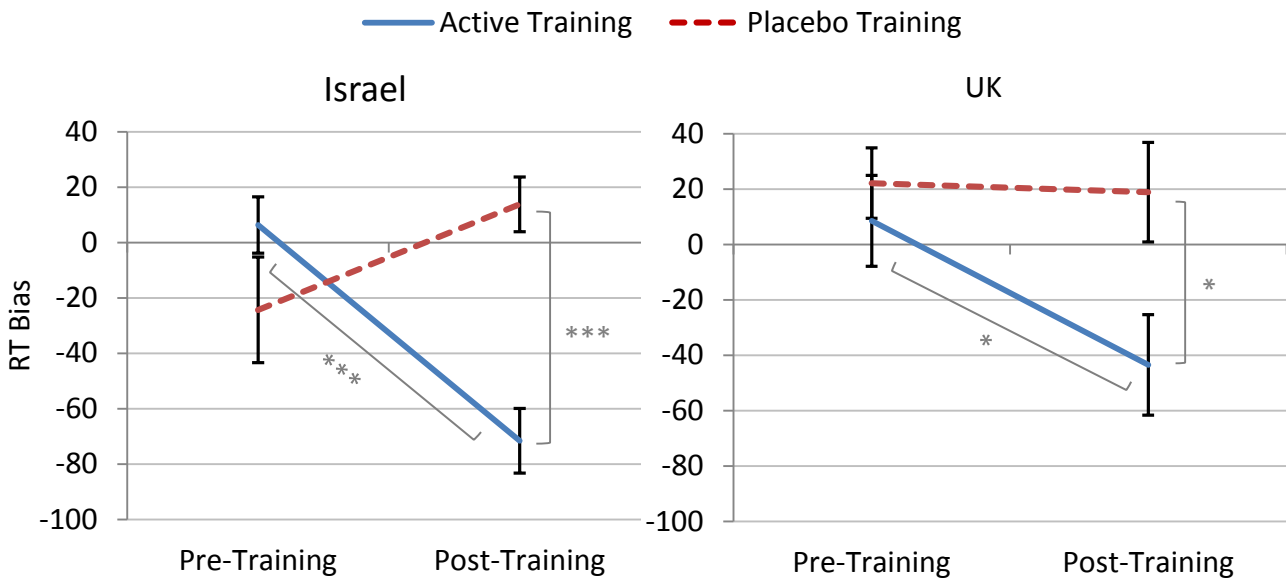


Figure 2: Mean RT bias scores at pre- and post-training for the active and placebo training groups in each site. Error bars represent standard errors.

Generalization of interpretation bias modification

Percent of anger responses: The ANOVA yielded main effects of Time, $F(1, 148) = 21.06, p < .001, \eta^2_p = .13$, and Training Condition $F(1, 148) = 26.57, p < .001, \eta^2_p = .15$, qualified by a significant Time $\times$ Training Condition interaction $F(1, 148) = 44.56, p < .001, \eta^2_p = .23$ (for means and CIs see Table 1). There was also a Time $\times$ Training Condition $\times$

Task Version interaction effect, $F(1, 148) = 3.99$, $p = .047$, $\eta^2_p = .03$. This suggests a somewhat different pattern of generalization between the two task versions. However, as can be seen in Figure 3, percent of anger responses reduced from pre- to post-training in the active training groups of both task versions, $t(38) = 7.41$, $p < 0.001$, Cohen's $d = 1.20$ and $t(37) = 3.85$, $p < .001$, Cohen's $d = .64$ for the TAU and UoB task versions, respectively. In contrast, in the placebo training groups of both task versions percent of anger responses increased from pre- to post-training, but there was only weak statistical evidence for this ($t_s < 1.75$ $p_s > .088$). Moreover, for both task versions (TAU, UoB) the two training conditions (active, placebo) differed in percent of anger responses after training, $t(77) = 7.14$, $p < .001$ and $t(75) = 3.12$, $p = .003$ for TAU and UoB task versions, respectively, but there was no statistical evidence for this before training ($t_s < .66$, $p_s > .250$).

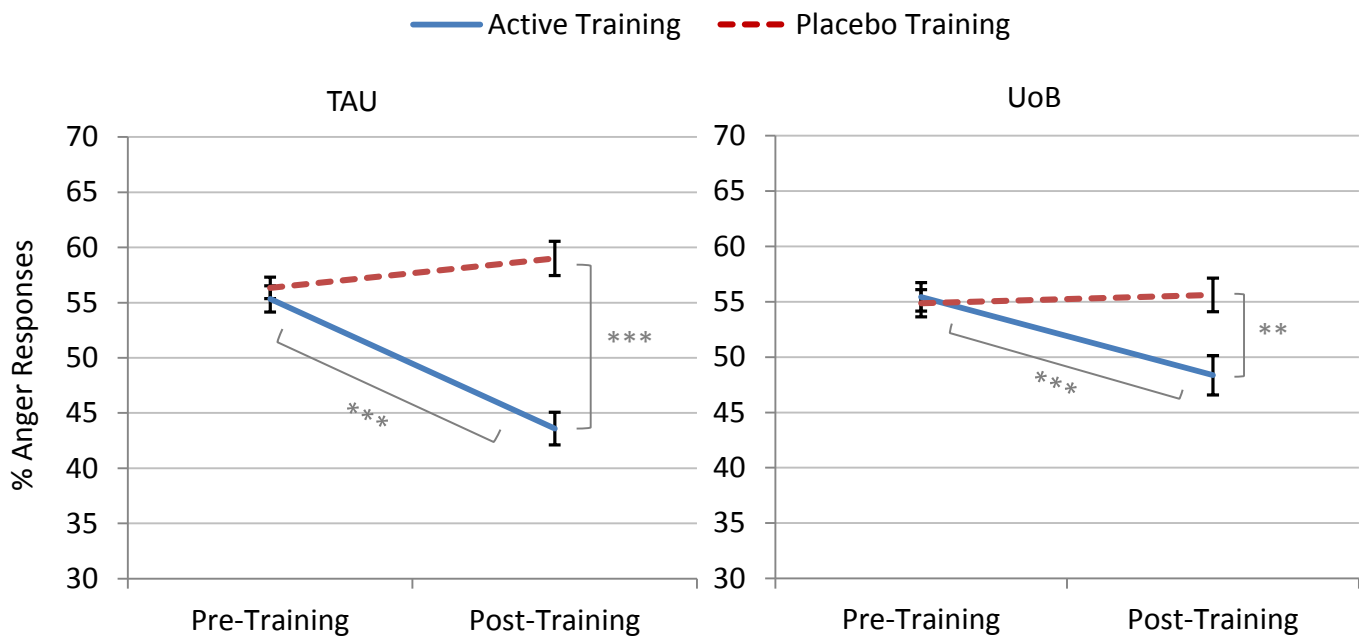


Figure 3: Mean percent of anger responses in generalization block at pre- and post-training for the active and placebo training groups in each task version. Error bars represent standard errors.

RT bias scores: The ANOVA yielded a main effect of Training Condition $F(1, 148) = 9.69, p = .002, \eta^2_p = .06$, qualified by a Time $\times$ Training Condition interaction $F(1, 148) = 6.22, p = .014, \eta^2_p = .04$ (for means and CIs see Table 1). This interaction did not appear to be moderated by Task Version or Site (Time $\times$ Training Condition $\times$ Site and Time $\times$ Training Condition $\times$ Task Version interactions, $F_s < 1.50, p_s > .223$).

Ultimatum Game

Baseline proposal: The ANOVA yielded no effects of training condition, task version, or site, or any interactions ($F_s < 1.46, p_s > .228$). Mean baseline proposals in both training conditions were close to a fair split of the sum ($M = 49.6\%$, 95% CI = [46.6%, 52.6%] and $M = 48.6\%$, 95% CI = [46.5%, 50.7%] for active and placebo training groups, respectively).

Direct retaliation: There was no evidence that rejection rates of unfair offers (round 2 in the Ultimatum Game) differed between the active (61%) and placebo (62%) groups, $\chi^2 < .02, p > .250$. The ANOVA on percent of the sum offered to the "unfair" player (round 4) yielded no evidence of an effect for Training Condition, Task Version, or Site, nor for any interaction effects ($F_s < 1.46, p_s > .228$). Notably, mean proposals to the "unfair" player in both training conditions were lower than half of the sum ($M = 40.1\%$, 95% CI = [36.0%, 44.3%] and $M = 37.3\%$, 95% CI = [34.1%, 40.6%] for active and placebo training groups, respectively), indicating that despite lack of group differences the anger manipulation (i.e., receiving an unfair offer) resulted in direct retaliatory behavior toward the "unfair" player.

Displaced retaliation: The ANOVA yielded a main effect of Training Condition, $F(1, 148) = 3.93, p = .049, \eta^2_p = .03$, indicating that on average participants in the active training group tended to give fairer offers to the "innocent" player ($M = 48.3\%$, 95% CI = [45.0%, 51.6%]) relative to the placebo training group ($M = 43.9\%$, 95% CI = [40.9%, 46.9%]),

(Figure 4). There was no clear evidence for other main or interaction effects ($F_s < 1.81$, $p_s > .181$).

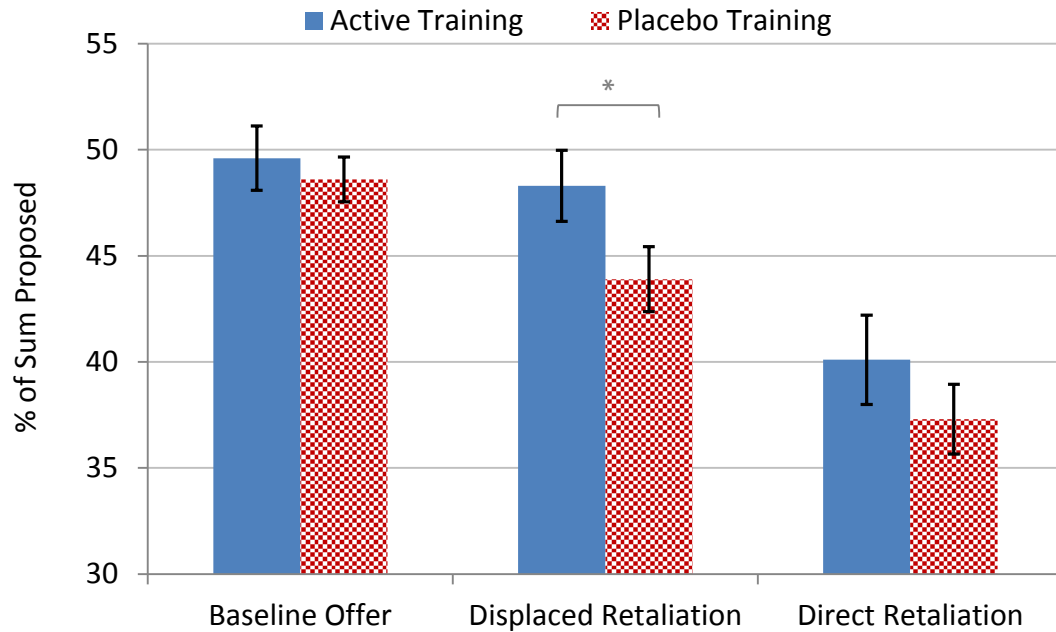


Figure 4: Mean percent of the sum proposed by participants in the active and placebo training groups in each round of the Ultimatum Game. Error bars represent standard errors.

Change in anger mood during the sessions

Only a main effect of Training Condition was found, $F(1, 147) = 5.61$, $p = .019$, $\eta^2_p = .04$, indicating that participants in the placebo training condition tended to get more angry during the sessions (mean anger change = 1.66, 95% CI = [0.74, 2.58]) compared to participants in the active training group (mean anger change = 0.09, 95% CI = [-0.85, 1.02]).

Self-reported trait anger scores

No effect of Time, Training Condition or their interaction on trait anger scores or anger expression index were found, $F_s < 1.08$, $p_s > .250$ (for means and CIs see Table 1). A main effect of Site was found for Trait Anger, $F(1, 148) = 143.96$, $p < .001$, $\eta^2_p = .49$, and for

the Anger Expression Index, $F(1, 148) = 66.09$, $p < .001$, $\eta^2_p = .31$, which was expected due to the different cut off scores used for screening in the two samples, resulting in higher mean trait anger and anger expression scores in the UK sample ($M = 28.75$, $CI = [27.86, 29.64]$ and $M = 54.40$, $CI = [51.93, 56.87]$, respectively) compared to the Israel sample ($M = 21.23$, 95% $CI = [20.36, 22.09]$ and $M = 40.23$, 95% $CI = [37.83, 42.63]$, respectively).

Discussion

We explored the effect of two versions of cognitive training on interpretation of ambiguous faces, retaliatory behavior, and self-reported anger in two samples of undergraduate students with high levels of trait anger. Three main findings emerged. First, training modified interpretations of ambiguous faces; specifically, relative to placebo training active training reduced perception of anger and speeded perceptions of happiness. Importantly, these effects generalized to novel ambiguous faces, not used during training (Dalili et al., 2016; Griffiths et al., 2015). Second, training changed patterns of retaliatory behavior on the Ultimatum Game; specifically, subjects who received active training were less likely than those who received placebo training to engage in displaced retaliatory behavior. Finally, training did not change rated levels of trait anger, but mitigated elevations in state anger during the experimental sessions.

We expected training to reduce both direct and displaced retaliation in the Ultimatum Game. As expected, active compared to placebo training reduced displaced retaliation, as reflected in more fair offers to the "innocent" player. However, unexpectedly, groups did not differ in rates of direct retaliation, as reflected by offers to the "unfair" player. This could reflect more intense negative feelings toward "unfair" relative to "innocent" players, generating affect-driven behaviors that are less malleable through the current training protocols. Future studies could examine whether more training sessions or more trials per

session could prove more efficient in modifying direct retaliation. Another possibility is that anger toward the "unfair" player was more conscious and explicit relative to the displaced anger situation and thus involved more elaborate cognitive top-down processes relative to the displaced anger condition. For example, since participants in the current study did not know how many rounds they would play with each of the players, their retaliation may not have been solely emotional, but rather a strategy designed to "educate" the selfish player for potential future rounds of the game. In contrast, displacement of anger toward the "innocent" player may have been less conscious, and less controlled by explicit strategy, and therefore more readily amenable to implicit modification. Prior research suggests that emotion regulation via cognitive reappraisal reduces the impact of anger on decision making during the Ultimatum Game (Fabiansson and Denson, 2012). This previous finding corresponds with the current findings given that cognitive reappraisal, like the current computerized training, is based on promoting more positive interpretations. However, while the cognitive training used here reduced only displaced retaliation, cognitive reappraisal lead to more fair offers in general (both to the provoking counterpart and to non-provoking counterparts). This difference may be related to the fact that cognitive reappraisal is based on elaborative top-down thinking processes and may therefore have more impact on conscious direct anger toward the provocateur.

In the current study cognitive training did not affect self-reported trait anger. This may be related to the fact that pre- and post-training trait anger were measured only one week apart. The trait anger questionnaire is based on ratings regarding "usual behavior". Thus, it may not be sufficiently sensitive to measure short-term changes such as the ones we attempted to detect in the current study. Future research may try to measure changes in trait anger following training over longer time periods, or alternatively be more specific when evaluating change in anger levels (e.g., ask specifically about the previous week). Future

studies could also consider use of other measures, such as partner reports regarding change in anger levels, rather than relying solely on self-reports.

Our study has the advantage of basing its findings on a large sample collected across two different sites. Yet, the current findings should also be considered in light of several limitations. First, the anger-provoking manipulation used in the current study was receiving an unfair offer in the ultimatum game. While it is well established that unfair offers in this game evoke anger (Pillutla and Murnighan, 1996), this may be considered to be a mild and context-specific provocation. Future studies could use stronger provocations such as direct negative feedback (e.g., Reijntjes et al., 2013) or harassment (e.g., Lobbestael et al., 2008) to test the impact of training. Second, due to an error, the samples were screened based on different cut off scores. Thus, the samples differed significantly in trait anger levels with above average trait anger levels in the Israeli site and extremely high trait anger levels in the UK site. However, the fact that the training effects were similar in both sites may suggest that such training could effectively impact interpretation patterns and displaced retaliatory behavior across a wide range of trait anger intensity. Third, it is important to keep in mind that despite their rather high trait anger levels, all participants in the current study were undergraduate students who may reflect a rather functional and well-behaved population. More research is needed to explore the behavioral effects of these training interventions on other populations that show maladaptive functioning due to anger control problems. Such an effect has been demonstrated by Penton-Voak et al. (2013), who reported that training resulted in fewer instances of aggression among youth at risk for criminal behavior. However, replications are needed, as well as better understanding regarding the differential effect of this training on direct and displaced anger-related behaviors.

In conclusion, our results suggest that decisions and behaviors related to displaced anger may be diminished via modification of anger-related interpretation processes. This

finding may have important implications in the context of personal relationships as well as in the context of economic trading and negotiations. This encouraging prospect should be empirically grounded by future randomized controlled trials.

References

- Adler, A.B., Britt, T.W., Castro, C.A., McGurk, D., Bliese, P.D., 2011. Effect of transition home from combat on risk-taking and health-related behaviors. *Journal of Traumatic Stress* 24 (4), 381-389.
- Andrade, E.B., Ariely, D., 2009. The enduring impact of transient emotions on decision making. *Organizational Behavior and Human Decision Processes* 109 (1), 1-8.
- Ben-Shakhar, G., Bornstein, G., Hopfensitz, A., van Winden, F., 2007. Reciprocity and emotions in bargaining using physiological and self-report measures. *Journal of economic psychology* 28 (3), 314-323.
- Crick, N.R., Dodge, K.A., 1994. A review and reformulation of social information-processing mechanisms in childrens social-adjustment. *Psychological Bulletin* 115 (1), 74-101.
- Dalili, M.N., Schofield-Toloz, L., Munafò, M.R., Penton-Voak, I.S., 2016. Emotion recognition training using composite faces generalises across identities but not all emotions. *Cognition and Emotion*, 1-10.
- Dodge, K.A., 2006. Translational science in action: Hostile attributional style and the development of aggressive behavior problems. *Development and Psychopathology* 18 (3), 791-814.
- Dollard, J., Miller, N.E., Doob, L.W., Mowrer, O.H., Sears, R.R., 1939. Frustration and aggression. Yale University Press, New Haven, CT.
- Fabiansson, E.C., Denson, T.F., 2012. The Effects of Intrapersonal Anger and Its Regulation in Economic Bargaining. *PloS one* 7 (12), e51595.
- Freud, A., 1937. The ego and the mechanisms of defence. Hogarth, London.

- Griffiths, S., Jarrold, C., Penton-Voak, I.S., Munafò, M.R., 2015. Feedback training induces a bias for detecting happiness or fear in facial expressions that generalises to a novel task. *Psychiatry Research* 230 (3), 951-957.
- Güth, W., Schmittberger, R., Schwarze, B., 1982. An experimental analysis of ultimatum bargaining. *Journal of economic behavior & organization* 3 (4), 367-388.
- Harlé, K.M., Sanfey, A.G., 2007. Incidental sadness biases social economic decisions in the Ultimatum Game. *Emotion* 7 (4), 876.
- Hoobler, J.M., Brass, D.J., 2006. Abusive supervision and family undermining as displaced aggression. *Journal of Applied Psychology* 91 (5), 1125.
- Litvak, P.M., Lerner, J.S., Tiedens, L.Z., Shonk, K., 2010. Fuel in the fire: How anger impacts judgment and decision-making, *International handbook of anger*. Springer, pp. 287-310.
- Lobbestael, J., Arntz, A., Wiers, R.W., 2008. How to push someone's buttons: A comparison of four anger-induction methods. *Cognition & Emotion* 22 (2), 353-373.
- Lundqvist, D., Flykt, A., Öhman, A., 1998. The Karolinska directed emotional faces (KDEF). CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, 91-630.
- Maoz, K., Adler, A.B., Bliese, P.D., Sipos, M.L., Quartana, P.J., Bar-Haim, Y., 2016a. Attention and interpretation processes and trait anger experience, expression, and control. *Cognition and Emotion*, 1-12.
- Maoz, K., Eldar, S., Stoddard, J., Pine, D.S., Leibenluft, E., Bar-Haim, Y., 2016b. Angry-happy interpretations of ambiguous faces in social anxiety disorder. *Psychiatry Research* 241, 122-127.

- Marcus-Newhall, A., Pedersen, W.C., Carlson, M., Miller, N., 2000. Displaced aggression is alive and well: a meta-analytic review. *Journal of Personality and Social Psychology* 78 (4), 670.
- Orobio de Castro, B., 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 (3), 916-934.
- Owen, J.M., 2011. Transdiagnostic cognitive processes in high trait anger. *Clinical Psychology Review* 31 (2), 193-202.
- Penton-Voak, I.S., Thomas, J., Gage, S.H., McMurrin, M., McDonald, S., Munafò, M.R., 2013. Increasing recognition of happiness in ambiguous facial expressions reduces anger and aggressive behavior. *Psychological Science* 24 (5), 688-697.
- Pillutla, M.M., Murnighan, J.K., 1996. Unfairness, anger, and spite: Emotional rejections of ultimatum offers. *Organizational Behavior and Human Decision Processes*.
- Plutchik, R., 2001. The nature of emotions. *American Scientist* 89 (4), 344-350.
- Psychology Software Tools, Inc. [E-Prime 2.0]. (2012). Retrieved from <http://www.pstnet.com>.
- Reijntjes, A., Kamphuis, J.H., Thomaes, S., Bushman, B.J., Telch, M.J., 2013. Too calloused to care: An experimental examination of factors influencing youths' displaced aggression against their peers. *Journal of experimental psychology: general* 142 (1), 28.
- Sanfey, A.G., Rilling, J.K., Aronson, J.A., Nystrom, L.E., Cohen, J.D., 2003. The neural basis of economic decision-making in the ultimatum game. *Science* 300 (5626), 1755-1758.
- Sapientfield, B.R., 1954. Displacement, Personality dynamics: An integrative psychology of adjustment. Alfred A. Knopf, New York, pp. 309-326.

- Scherer, K.R., Wranik, T., Sangsue, J., Tran, V., Scherer, U., 2004. Emotions in everyday life: probability of occurrence, risk factors, appraisal and reaction patterns. *Social Science Information Sur Les Sciences Sociales* 43 (4), 499-570.
- Schultz, D., Grodack, A., Izard, C., E, 2010. State and trait anger, fear, and social information processing, in: Potegal, M., Stemmler, G., Spielberger, C. (Eds.), *International Handbook of Anger: Constituent and Concomitant Biological, Psychological and Social Processes*. Springer, pp. 311-325.
- Spielberger, C.D., 1999. Professional manual for the state-trait anger expression inventory-2 (STAXI-2). Psychological Assessment Resources, Inc., Odessa, FL.
- Stoddard, J., Sharif-Askary, B., Harkins, E.A., Frank, H.R., Brotman, M.A., Penton-Voak, I.S., Maoz, K., Bar-Haim, Y., Munafo, M., Pine, D.S., 2016. An Open Pilot Study of Training Hostile Interpretation Bias to Treat Disruptive Mood Dysregulation Disorder. *Journal of Child and Adolescent Psychopharmacology*.
- Tiddeman, B., Burt, M., Perrett, D., 2001. Prototyping and transforming facial textures for perception research. *IEEE computer graphics and applications* 21 (5), 42-50.
- Tottenham, N., Tanaka, J.W., Leon, A.C., McCarry, T., Nurse, M., Hare, T.A., Marcus, D.J., Westerlund, A., Casey, B.J., Nelson, C., 2009. The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research* 168, 242-249.
- 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 (1), 3-21.
- Wilkowski, B.M., Robinson, M.D., 2010. The anatomy of anger: an integrative cognitive model of trait anger and reactive aggression. *Journal of Personality* 78 (1), 9-38.
- Wranik, T., Scherer, K.R., 2010. Why do I get angry? a componential appraisal approach, in: Potegal, M., Stemmler, G., Spielberger, C. (Eds.), *International Handbook of Anger:*

Constituent and Concomitant Biological, Psychological and Social Processes.
Springer, pp. 243-266.

Figure 1: Mean percent of anger responses at pre- and post-training for the active and placebo training groups. Error bars represent standard errors.

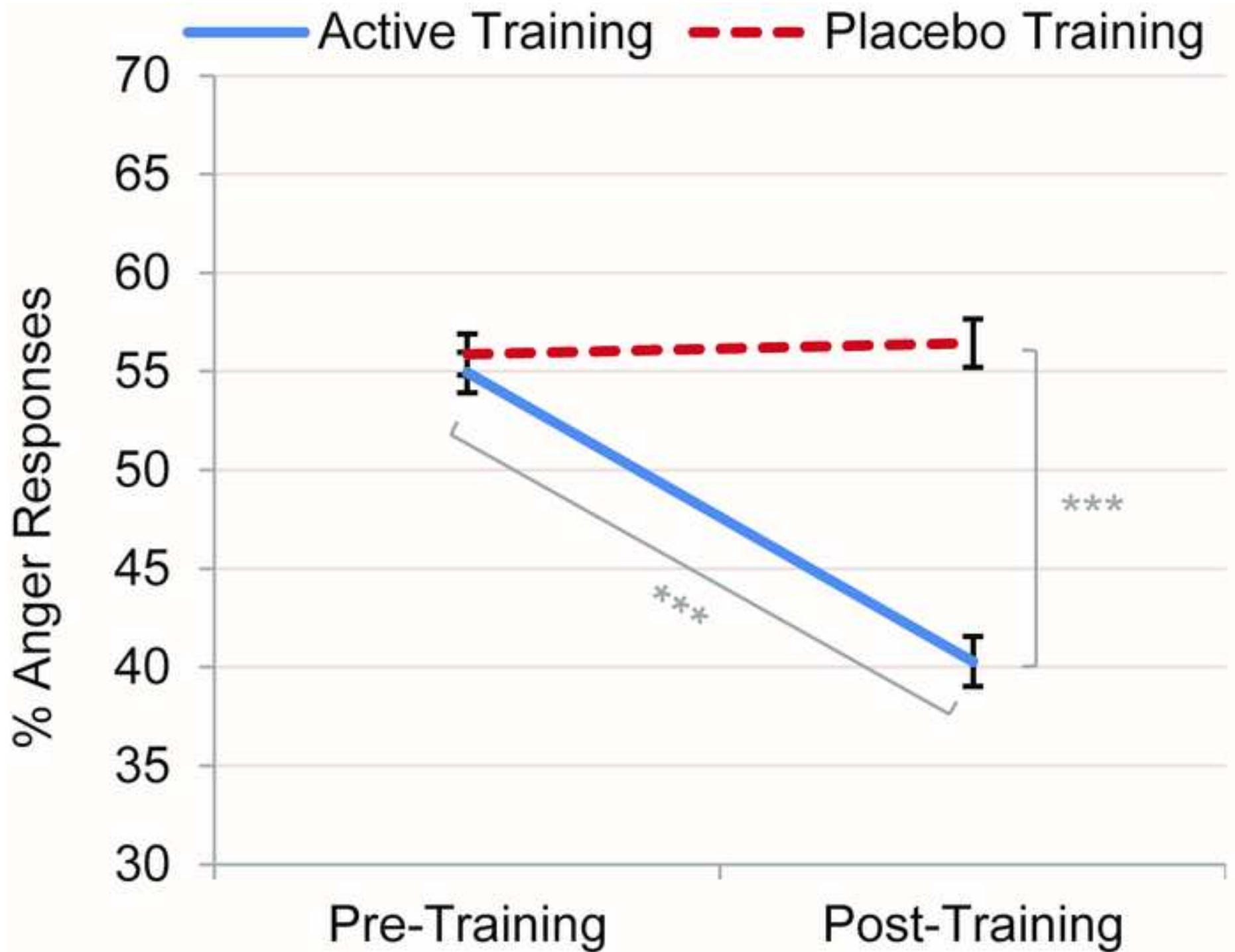


Figure 2: Mean RT bias scores at pre- and post-training for the active and placebo training groups in each site. Error bars represent standard errors.

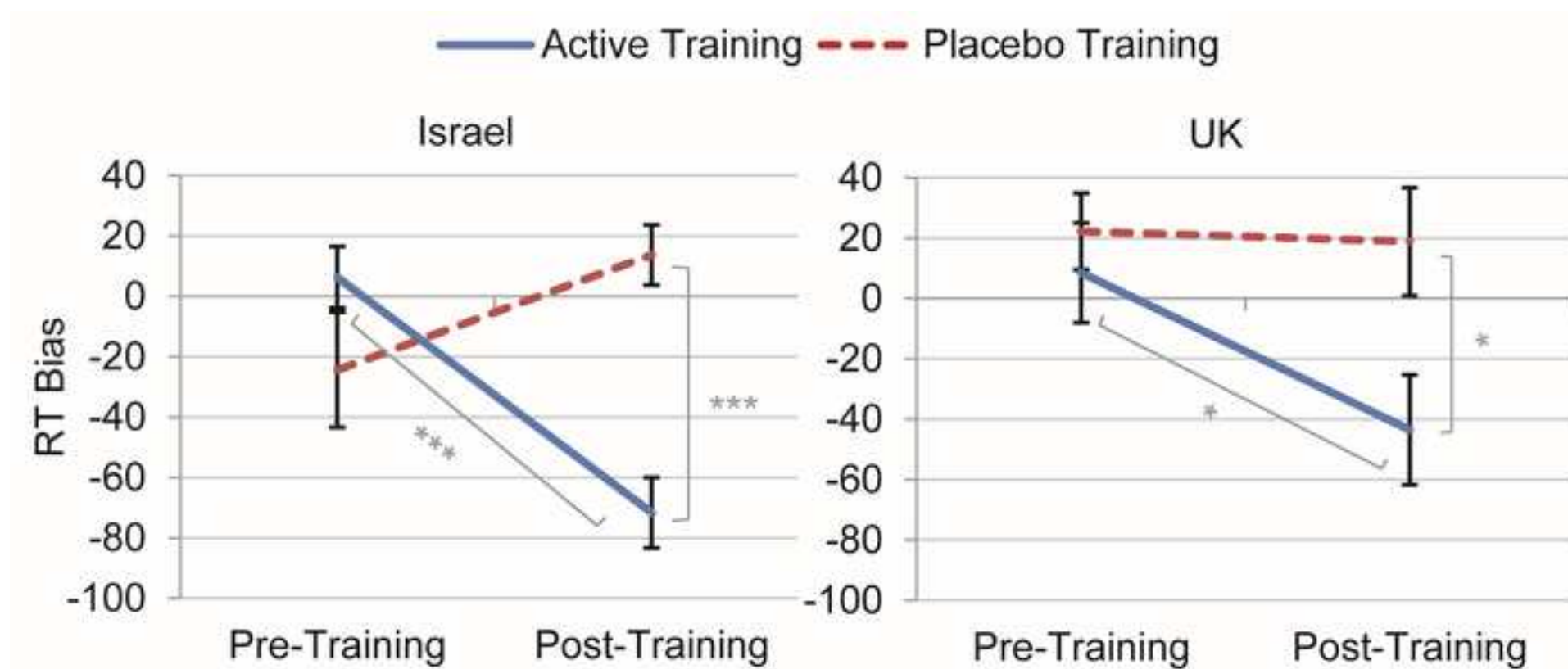


Figure 3: Mean percent of anger responses in generalization block at pre- and post-training for the active and placebo training groups in each task version. Error bars

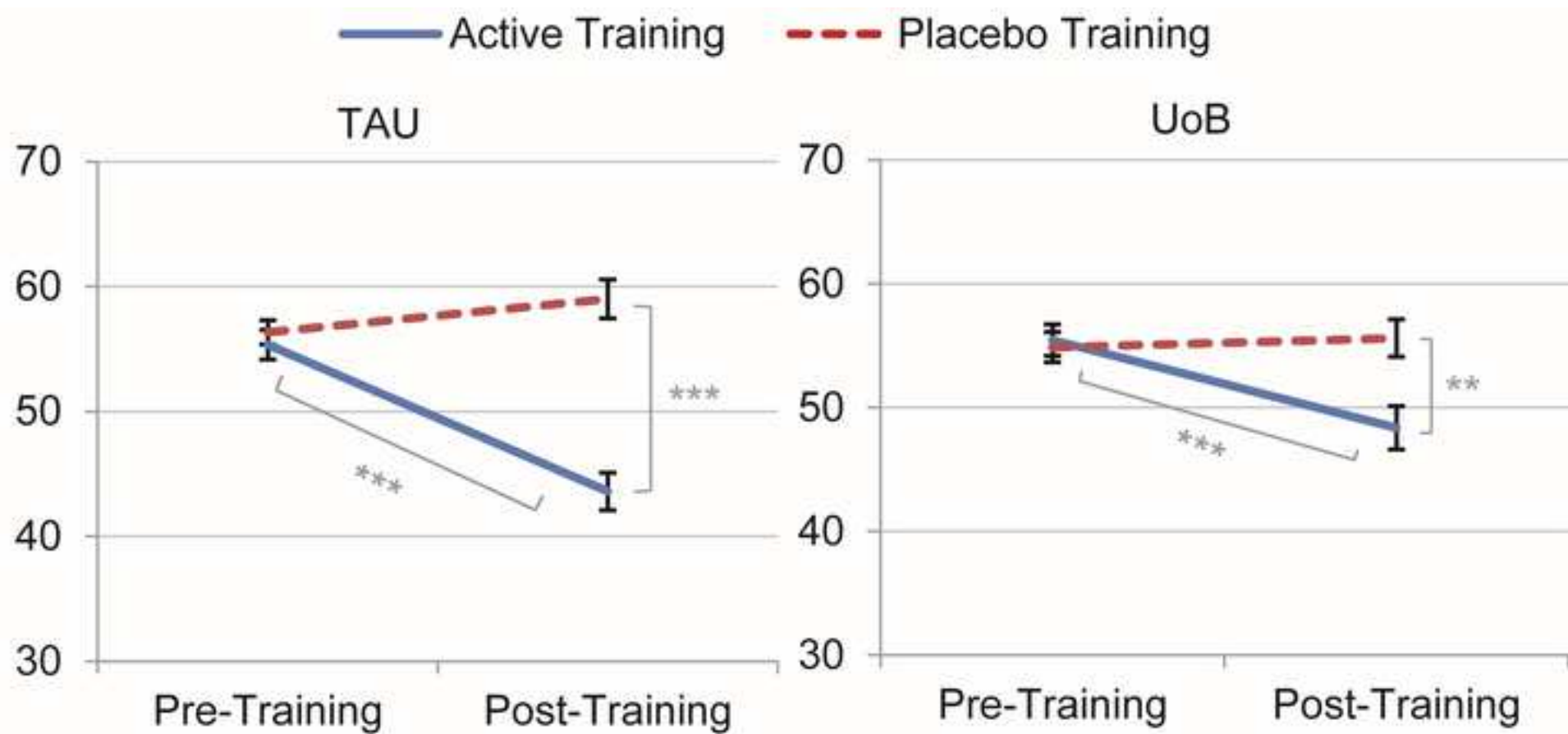


Figure 4: Mean percent of the sum proposed by participants in the active and placebo training groups in each round of the Ultimatum Game. Error bars represent standard

