Interpretation bias modification for hostility

American Psychological Association

Journal of Consulting and Clinical Psychology

ISSN: 0022-006X

2021, Vol. 89, No. 5, 421–434
https://doi.org/10.1037/ccp0000651

Interpretation Bias Modification for Hostility: A Randomized Clinical Trial

Martijn W. van Teffelen[1, 2], Jill Lobbestael[1], Marisol J. Voncken[1], Jesse R. Cougle[3], and Frenk Peeters[1]

[1] Department of Clinical Psychological Science, Maastricht University

[2] Department of Anxiety Disorders, METggz

[3] Department of Psychology, Florida State University

Objective: Hostility is a transdiagnostic phenomenon that can have a profound negative impact on interpersonal functioning and psychopathological severity. Evidence suggests that cognitive bias modification for interpretation bias (CBM-I) potentially reduces hostility. However, stringent efficacy studies in people with clinical levels of hostility are currently lacking. ***Method:*** The present study investigated the effects of CBM-I in two studies: one feasibility study (Study 1) in a mixed clinical-community sample of men ($N = 29$), and one randomized clinical study (Study 2) in a mixed-gender sample with clinical levels of hostility ($N = 135$), pre-registered at https://osf.io/r46jn. We expected that CBM-I would relate to a larger increase in benign interpretation bias and larger reductions in hostile interpretation bias, hostility symptoms and traits, and general psychiatric symptoms at post-intervention compared to an active control (AC) condition. We also explored the beneficial carry-over effects of CBM-I on working alliance in subsequent psychotherapy 5 weeks after finishing CBM-I ($n = 17$). ***Results:*** Results showed that CBM-I increased benign interpretation bias in both studies and partially reduced hostile interpretation bias in Study 2, but not in Study 1. Findings of Study 2 also showed greater reductions in behavioral (but not self-reported) aggression in CBM-I relative to control, but no condition differences were found in self-report hostility measures and general psychiatric symptoms. ***Conclusions:*** Overall, we found modest support for CBM-I as an intervention for hostility, with some evidence of its efficacy for hostile interpretation bias and aggression. We discuss study limitations as well as directions for future research.

What is the public health significance of this article?
Interpreting ambiguous everyday situations in a hostile way is common for people experiencing hostility. Overall, we found modest support for interpretation training as an intervention for hostility, with some evidence of its efficacy for hostile interpretation bias and aggression. We discuss study limitations as well as directions for future research.

Keywords: cognitive bias modification for interpretation bias, hostile interpretation bias, hostility, anger, aggression

Supplemental materials: https://doi.org/10.1037/ccp0000651.supp

Hostility is a trait constellation consisting of a tendency to hold a hostile attributional style, experience angry affect, and behave aggressively, and is considered a trans-diagnostic clinical phenomenon (Cassiello-Robbins & Barlow, 2016). Clear prevalence estimates in clinical samples are currently missing, however one observational study in 3,800 outpatients indicated that 43.60% reported moderate to severe anger and 21.20% reported moderate to severe aggressive behavior in the preceding week (Genovese et al., 2017). Next to a profound impact on negative interpersonal functioning (Henrichs et al., 2015), hostility is associated with increased psychopathological severity (Cassiello-Robbins & Barlow, 2016) and suicidality (Ammerman et al., 2015).

Treatment options for hostility exist. However, effects appear less pronounced than those for other psychopathologies (e.g., panic disorder and body dysmorphic disorder), as 34% of patients do not profit from treatment (Hofmann et al., 2012). Moreover, premature treatment discontinuation in patients with hostility is significant (Arntz et al., 2015; Cassiello-Robbins et al., 2015; Putt et al., 2001) and few high-quality treatment effects studies on hostility have been conducted (Del Vecchio & O'Leary, 2004). Clinically, patients with increased levels of hostility are often described by therapists as "challenging" (von der Lippe et al., 2008).

One potential promising novel way of reducing hostility is offered by cognitive bias modification for interpretation bias (CBM-I).

Martijn W. van Teffelen https://orcid.org/0000-0003-2411-5323
Marisol J. Voncken https://orcid.org/0000-0001-8521-3109

We have no conflicts of interest to disclose. This study was preregistered at osf.io/na9rs. We thank Dieuwertje Abeling-Boselie, Renée Smeets, Loucky Mulder, Lisette Schlangen, Astrid Dello, Ronald Cronenburg, Anne Tiemensma, Eric Munsters, Michiel Vestjens, Ellen Huben, Maya Molag, Ramona Waelen, Linda Litjens, Ingrid Weijnen, Miriam Schoutrop, Rosy Ghijsens and clinical and policy staff at METggz, U-Center Epen and Mondriaan Zorggroep for their contribution to this work.

Correspondence concerning this article should be addressed to Martijn W. van Teffelen, Department of Clinical Psychological Science, Maastricht University, P.O. Box 616, 6200 MD, Maastricht, the Netherlands. Email: martijn.vanteffelen@maastrichtuniversity.nl

CBM-I is a computerized procedure that targets an important aspect of hostility, that is, hostile interpretation bias, referring to a tendency to interpret emotionally ambiguous scenarios in a hostile way. This is achieved by presenting patients with many (unfamiliar) emotionally ambiguous scenarios on a computer followed by a reinforcement of benign instead of threatening interpretations (Mathews & Mackintosh, 2000). In the present work, we aim to investigate the effects of CBM-I on hostile interpretation bias in one randomized sham-controlled feasibility study and one randomized sham-controlled clinical study.

Clinically, CBM-I was first developed and tested in participants with anxious and depressive symptomatology (e.g., Beard & Amir, 2008; Bowler et al., 2012). Meta-analytic evidence supported the efficacy of CBM-I on anxiety- and depression-related interpretation biases with a pooled effect size of $g = .81$ (Hallion & Ruscio, 2011). More recently, a meta-analysis on the efficacy of CBM-I on anxiety and depressive symptomatology showed only small standardized mean differences (SMD) of SMD $= -.30$ for anxiety symptoms and SMD $= -.26$ for depressive symptoms (Fodor et al., 2020). However, inconsistency of findings (specifically for depression studies), heterogeneity, and risk of bias potentially impede reliable interpretation of these findings. Thus, at present it is unclear whether CBM-I is efficacious at all, and how large the effect is. Moreover, these meta-analytic studies on the effects of CBM-I in anxiety and depression did not include hostile interpretation bias as an outcome.

Preliminary evidence supporting the efficacy of CBM-I on hostile interpretation bias is provided by five studies. A first study showed that hostile and benign interpretation bias can be experimentally manipulated (following one CBM-I session of 64 different scenario-training trials) in both benign and hostile directions in students (Hawkins & Cougle, 2013). This study also showed that participants experienced less anger when confronted with a provocative insult following benign interpretation training. The second study showed that three sessions (i.e., 15 non-computerized scenario-training trials per session) of benign interpretation training in a relatively small sample of aggressive boys were related to greater improvements of benign and hostile interpretation bias, anger, and self-reported aggression compared to untrained controls (Vassilopoulos et al., 2015). A third randomized sham-controlled study in non-treatment seeking people with alcohol use disorder and elevated levels of trait anger showed that eight sessions of CBM-I (i.e., 64 trials of scenario training per session) was related to greater improvements in benign and hostile interpretation bias, trait anger and self-reported anger expression (Cougle et al., 2017). The fourth study compared the effects of benign versus hostile training (i.e., one session of 52 imagery-scenarios training trials) in students, and showed that benign training was related to an increase in prosocial interpretations (i.e., an overall bias score calculated by subtracting mean negative from mean positive interpretation ratings) and a reduction in anger and self-reported verbal aggression, whereas hostile training did not relate to significant changes in these outcomes (AlMoghrabi et al., 2018). A fifth and final study in non-treatment seeking college students with major depressive disorder compared an eight-session CBM-I training (i.e., 64 scenario training trials per session) to a sham control condition and demonstrated that CBM-I was related to greater improvement in benign and hostile interpretation bias and anger control. However, no effects were found on depressive interpretation bias, depressive symptoms, or trait anger (Smith et al., 2018). Taken together, preliminary evidence suggests that CBM-I may be moderately efficacious in reducing hostile interpretation bias. Moreover, next to scenario-training, another technique that is often used in CBM-I research is the word-sentence association paradigm (WSAP; Beard & Amir, 2008). In this paradigm interpretation bias is trained by offering word-sentence pairs, of which the sentences are emotionally ambiguous and the words are either neutral or non-neutral (e.g., threatening). The WSAP technique has not been used before in CBM-I research for hostile interpretation bias. To the best of our knowledge, a methodologically stringent randomized sham-controlled clinical study in (adult) people with clinical levels of hostility is currently lacking.

Recent evidence also suggests that CBM-I may perhaps additionally augment the efficacy of existing therapy protocols. Two studies showed that offering CBM-I prior to (computerized) cognitive behavior therapy enhanced training effects on anxiety symptoms compared to sham-training prior to cognitive behavior therapy (Beard et al., 2019; Butler et al., 2015). The idea of offering CBM-I prior to existing therapy protocols may be even more attractive for populations with increased levels of hostility. That is, patients with increased hostility levels are at an increased risk to engage in hostile interactions with therapists (von der Lippe et al., 2008), which negatively impacts working alliance (Gülüm et al., 2018). One study showed that working alliance positively mediated the relationship between low levels of hostility and treatment outcome in dialectical behavior therapy for borderline personality disorder (Hirsh et al., 2012). Offering CBM-I prior to psychotherapy may therefore have beneficial effects on psychotherapy on top of its general effects, through increased working alliance levels.

Taken together, preliminary evidence suggests that CBM-I potentially reduces hostile interpretation bias, anger, self-reported aggression, and increases benign interpretation bias and working alliance. In the present work, we tested an eight-session CBM-I intervention (i.e., consisting of both scenario- and WSAP training) and compared its effects to an active control condition in two studies. We tested whether our CBM-I intervention would alter hostile interpretation bias by increasing benign interpretation bias and reducing hostile interpretation bias. We also examined the effects of CBM-I on state and trait hostility (i.e., angry affect, hostile attitudes, and aggressive behavior), and general psychiatric symptoms, and tested whether CBM-I affected working alliance in subsequent psychotherapy. The first study (Study 1) served as a feasibility study to establish whether CBM-I altered interpretation biases (the main outcomes) and state anger in the desired direction. It was followed by a randomized sham-controlled clinical study (RCT; Study 2) in people with clinical levels of hostility in which the efficacy of CBM-I on hostile interpretation bias, benign interpretation bias, and hostility outcomes (i.e., state and trait hostile thoughts, anger, and aggression) was assessed. The primary hypothesis was that CBM-I would result in a greater increase in benign interpretations and a greater reduction in hostile interpretation bias compared to active control training (AC) in both studies. The secondary hypothesis in Study 2 was that CBM-I would lead to greater reductions in hostility and general psychiatric symptoms. We also expected greater reductions in state-anger in both studies. Lastly, we explored the carry-over effects in Study 2 of CBM-I on subsequent psychotherapy and expected increased quality of working alliance with participants' therapists.

Study 1

Method

Participants

Participants were sampled in two ways. First, we recruited via flyers on university campus. Second, to ensure sufficient variation at the extreme end of the hostility dimension, we additionally recruited participants with clinically relevant levels of hostility (i.e., reported hostility complaints or showed aggressive behavior during diagnostic workup, yes or no) who were on a waiting list for treatment at two outpatient mental healthcare facilities in Maastricht, the Netherlands (i.e., METggz and Mondriaan Zorggroep). Only men were included, to reduce sample heterogeneity. Exclusion criteria were a waiting list shorter than 4 weeks,[1] age below 18 and above 60 years and illiteracy. Clinically relevant levels of hostility were assessed by clinical judgment of their therapist. Patients were excluded from participation, also based on clinical judgment, if they showed signs of current psychosis or mania, alcohol or drug abuse/dependency, and acute suicide risk. The total sample consisted of $N = 29$ participants (i.e., $n = 11$ patients and $n = 18$ students), of which $n = 16$ were in the CBM-I condition and $n = 13$ in the AC condition. Follow-up data were missing for $n = 9$ people. Mean age was $M = 40.1$ ($SD = 9.6$). The highest completed education level was low for $n = 4$, middle for $n = 12$, and high for $n = 13$ participants. Five participants were on active psychotropic medication. Chi-square and a t-test showed that age, educational level, and use of psychotropic medication did not significantly differ between conditions ($p > .071$).

Intervention Arms

CBM-I. This intervention condition consisted of eight sessions of benign intervention training, one session every 3 days. Prior to each session, participants engaged in an imagery task because a previous study reported that this increased the efficacy of CBM-I (Holmes et al., 2006). During this task, participants were instructed to imagine a lemon as vividly as possible on all sensory modalities with their eyes closed. This exercise prepared participants to imagine scenarios presented in the intervention as lively as possible. Each session consisted of two parts. In the first part, we developed and presented participants with 38 hostility related scenarios based on the WSAP (Beard & Amir, 2008). Here, participants were shown either a hostile (e.g., "rude") or neutral (e.g., "unaware") word, followed by an ambiguous sentence (e.g., "You are trying to concentrate, but someone is talking very loud"). Each ambiguous sentence was presented twice in a random order: once in combination with the hostile word and once in combination with the neutral word. Participants then had to indicate whether the word was related to the sentence by pressing "yes" or "no." Participants received positive feedback when they related neutral words to ambiguous sentences (i.e., a green "V"-sign accompanied by the text "Correct, the answer is yes/no"). Similarly, participants received negative feedback when they related hostile words to ambiguous sentences (i.e., "Incorrect, the answer should have been yes/no"). An example of the procedure is shown in Supplementary Figure 1. In the second part, participants were presented with 33 vignettes of the scenario training (Mathews & Mackintosh, 2000). The scenarios in this training were developed earlier (Cougle et al., 2017) and translated into Dutch for the current study. Here, participants were shown ambiguous sentences (e.g., "Someone near you laughs very loudly"). Then, participants had to complete a disambiguating word fragment (e.g., "This person is unaw_re of how loud he is"). After that, participants had to answer a comprehension question (e.g., "Is this person trying to annoy others?") by pressing "yes" or "no." Similar to the WSAP, participants received positive feedback when the comprehension question was answered in a benign way and negative feedback when the comprehension question was answered in a hostile way. For a graphical example, see Supplementary Figure 2.

Active Control Training. The active control condition was similar to the CBM-I condition, except that participants were presented with only neutral words during the WSAP (e.g., "ball" and "concentrate") and neutral scenarios (e.g., "Your friend is walking through the park," "He sees a squi_el" and "Did your friend see a bird?") during the scenario training.

Materials

Benign and Hostile Interpretation Bias. Interpretation tendencies were measured with the Social Information Processing-Attribution and Emotional Response questionnaire (SIP-AEQ; Coccaro et al., 2009). Participants were asked to read eight vignettes (e.g., "You tell a friend something personal and ask your friend not to discuss it with anyone else. However, a couple of weeks later, you find out that a lot of people know about it. You ask your friend why she/he told other people and your friend says: "Well, I don't know, it just came up and I didn't think it was a big deal."), followed by four statements (i.e., a direct hostile statement, indirect hostile statement, benign statement, and instrumental statement[2]). An example of a direct hostile statement is: "My friend wanted me to feel stupid for asking to keep my secret." Participants were then asked to indicate the likelihood of each statement on a 4-point Likert scale ranging from 0 (*not likely at all*) to 3 (*very likely*). The benign and hostile interpretation bias scores are calculated by respectively summing the likelihood scores of benign and hostile (i.e., direct and indirect) statements. Internal consistency of the SIP-AEQ is good (in Study 2 α's were .70 for benign interpretation bias and .90 for hostile interpretation bias scores) and convergent and discriminant validity are adequate (Coccaro et al., 2009).

State Anger. State anger was measured using the seven items of the anger subscale of the Profile Of Mood States (POMS; McNair et al., 1992). Participants were asked to indicate the extent to which items (i.e., "bad-tempered," "annoyed," "rebellious," "furious," "grouchy," "angry," and "on edge") reflect their current mood state on a 5-point Likert scale ranging from 0 (*not at all*) to 4 (*very much*). Internal consistency of the POMS is good and validity is adequate (Wicherts & Vorst, 2004).

Procedure

The Ethical Review Committee Psychology and Neuroscience at Maastricht University provided ethical approval to carry out the study

[1] Our CBM-I intervention took place over the course of 4 weeks. To avoid confounding effects of other therapies, people with a waiting time shorter than 4 weeks were excluded.

[2] Instrumental statements are not analyzed in the manuscript.

(ECP- 170_09_11_2014). The study took place between February and July 2016.[3] When participants arrived at the lab, the study's procedure was explained and written informed consent was obtained. In our information letter we explained participants that we were studying how we can reduce feelings of hostility with a computer training. We explained that our training aimed to change the thought processes that are involved in hostility. Next, participants were told that the experiment involved two experimental conditions; an active condition and a placebo condition to which participants were randomly allocated. An independent technician from another department at Maastricht University carried out permutated block (i.e., blocks of 4 and 6) randomization (stratified by gender) using https://www.randomizer.org/. The participants were blind to the condition. Participants completed benign and hostile interpretation bias and anger measures prior to engaging in the first session of their allocated intervention. The next seven sessions and post-intervention assessment were carried out at home. During the last session, benign and hostile interpretation bias and anger measures were again completed. The first and last session took approximately 1 hr, and the other sessions took about 20 min. At the end, participants were fully debriefed, thanked, and reimbursed with €35 for their participation.

Statistical Analyses

SPSS version 24 was used for all analyses. First, means and standard deviations were computed to examine the baseline characteristics of benign and hostile interpretation bias and anger measurements. Second, independent *t*-tests or Mann-Whitney *U*-tests were run to examine baseline differences. Third, to test the hypothesis that CBM-I predicts greater reductions in benign and hostile interpretation bias and anger three mixed regression models were run. Mixed regression was opted for giving its ability to handle missing data and modeling error terms, increasing statistical power (Baayen et al., 2008). In each model, condition, time, and condition by time interaction were entered as predictors. Within each regression model, repeated measures were clustered in participants. AR1 (first-order auto-regression) was selected as covariance structure, as −2 log likelihood testing showed that it was most parsimonious. Analyses were conducted following the intent-to-treat principle (i.e., including participants who dropped out). Missing data were handled using maximum likelihood estimation.

Results

Baseline levels of hostile interpretation bias, benign interpretation bias, and state anger showed no significant differences ($p > .139$) between the conditions, indicating that random allocation was successful. Study means and *SD*'s are shown in Table 1. In total, 15.52% of values were missing. Number of missed sessions did not differ per condition ($\chi^2 = 2.70$, $p = .101$) and Little's MCAR test indicated that they were missing completely at random ($\chi^2 = 5.56$, $p = .234$). Next, to test the hypotheses that CBM-I relates to a greater increase in benign interpretation bias, greater reduction in hostile interpretation bias, and greater reduction in state anger reactivity, three mixed regression models were run. First, findings on benign interpretation bias showed an effect of condition ($B = -3.80$, $SE = 1.54$, $t = -2.47$, $p = .020$), but not time ($B = -.20$, $SE = .69$, $t = -.28$, $p = .78$), that was qualified by a condition by time interaction ($B = 2.42$, $SE = 1.00$, $t = 2.41$, $p = .025$). Second, findings on hostile interpretation bias showed that the condition, time, and time by condition effects were not significant ($p > .245$). Third, results on state anger reactivity (i.e., change in anger from pre- to post-session) demonstrated that the time effect was significant ($B = 5.47$, $SE = 1.67$, $t = 3.28$, $p = .005$), whereas the effects of condition and the interaction were not significant ($p > .384$).

Study 1—Discussion

In Study 1, we examined the feasibility of an eight-session CBM-I intervention. The goal was to evaluate feasibility both in practical terms, but also in terms of changes in the target variables, benign and hostile interpretation bias. In line with our expectations, findings showed that CBM-I increased benign interpretation bias compared to AC. Contrary to our expectations, results showed no evidence of a decrease over time for hostile interpretation bias and anger reactivity for CBM-I, compared to AC. One explanation for this pattern of findings is that no such effects are present. Other explanations are that the current sample is too small to detect such effects, or that hostility levels were not high enough. In short, findings of Study 1 suggest that CBM-I has some feasibility, at least in terms of improvement in benign interpretation bias. Therefore, we tried to replicate these findings in a larger study, which was the focus of Study 2.

Table 1
Study 1—Descriptives

	CBM-I ($n = 16$)		AC ($n = 13$)	
Variable	Pre	Post	Pre	Post
SIP-AEQ hostile interpretation bias	2.61 (2.82)	1.44 (.88)	3.18 (2.56)	1.72 (2.41)
SIP-AEQ benign interpretation bias	5.00 (1.73)	7.22 (2.05)	6.36 (2.20)	6.18 (1.99)
State anger—Session 1 (POMS)[a]	7.89 (4.48)	4.22 (2.77)	6.72 (6.97)	2.18 (4.62)
State anger—Session 8 (POMS)[a]	1.44 (2.96)	.22 (.44)	2.18 (4.75)	2.36 (4.43)

Note. CBM-I = cognitive bias modification for interpretation; AC = active control condition; SIP-AEQ = social information processing-attribution and emotional processing questionnaire; POMS = Profile of Mood States.
[a] Non-normally distributed.

Study 2

Method

Participants

Participants were sampled in two ways. First, we recruited in two outpatient and one inpatient mental healthcare facility in the Maastricht, The Netherlands (i.e., METggz, Mondriaan Zorggroep, and U-Center). Target groups of these clinics include adult forensic and non-forensic patients with affective disorders, personality disorders, and addiction. Participants were screened for eligibility while on the waiting list for treatment. Second, participants were sampled in the local community using an advertisement in local news media asking for "people with a short fuse."[4] Inclusion criteria were aged between 18 and 60 years, mastery of the Dutch language, basic computer skills, and a score above 1.27 on the hostility scale of the Personality Inventory for DSM-5 (PID5-H; Van der Heijden et al., 2014).[5] This cutoff equals one standard deviation above the mean in both a Danish (a comparable population to the Netherlands) community as well as the mean of a clinical sample (Bach et al., 2016). Exclusion criteria were being on a waiting list for treatment shorter than 4 weeks,[6] IQ estimate below 80, observed psychotic/manic symptoms[7] during intake, suicidality, and not having access to a computer. A participant flow diagram is presented in Figure 1. In the pre-registration we determined sample size in the following way; to detect an effect of $d = 0.81$ (Hallion & Ruscio, 2011) with $\alpha = .05$ and $\beta = .10$, anticipating a 25% drop-out, $N1 = N2 = 10.5 * 2/0.81^2/.75 \approx 43$ participants are needed per condition. However, if we would have used the obtained effect size in S1 ($d = 0.51$)[8] to detect an effect with $\alpha = .05$ and $\beta = .20$ we would have needed a minimum of $n = 61$ participants per condition. In total, $N = 135$ people entered the study, of whom $n = 37$ were lost to follow-up. 62% of the sample consisted of people from the community. Sample characteristics are shown in Table 2.

Intervention Arms

The intervention arms were identical to Study 1 (i.e., combination of WSAP and scenario training).

Materials

Benign and Hostile Interpretation Bias. To measure benign and hostile interpretation bias we used the SIP-AEQ, similar to Study 1. In addition, we administered the Word Sentence Association Paradigm-Hostility (WSAP-H) scale (Dillon et al., 2016). In the WSAP-H, participants were presented with 16 hostile and 16 benign word-sentence pairs and were asked to indicated how well each word is related to the sentence on a 6-point Likert scale ranging from 1 (*not related at all*) to 6 (*very related*). Mean scores on the hostile and neutral word-sentence pairs are a measure for respectively hostile and benign interpretation bias. In the present study, we split up the WSAP-H into two parts to rule out memory effect and enable measurement of pre- and post-intervention. The instrument shows good internal consistency (in this study α range from .70 to .72) and adequate discriminant validity (Dillon et al., 2016).

State Measures. To measure different aspects of state hostility and general psychiatric symptoms in the past 3 days participants completed four measures. First, they were administered an adapted state-version of the eight-item hostility subscale of the Aggression Questionnaire (AQ-HS; Buss & Perry, 1992), measuring self-reported hostile thoughts in the past 3 days, which is scored on a 5-point Likert scale ranging from 1 (*totally disagree*) to 5 (*totally agree*; $\alpha = .90$). Second, we administered the 15-item State-Trait Anger Expression Inventory-2 state scale (STAXI-2S; Spielberger, 1999), measuring self-reported anger in the past 3 days, which is scored on a 4-point Likert scale ranging from 1 (*not at all*) to 4 (*very much*; $\alpha = .95$). Third, they completed an adapted state-version of the Forms of Aggression State Questionnaire (FOAS; Verona et al., 2008), measuring self-reported aggressive behavior in the past 3 days, which is scored on a 5-point Likert scale ranging from 1 ((*almost*) *never*) to 5 ((*almost*) *always*; $\alpha = .74$). In contrast to the original FOA, our participants were asked to indicate how often each behavior occurred in general in the past 3 days instead of "when angry." Moreover, we used an 11-item version that included the highest loading items from the original 40-item FOA in a separate sample of $N = 120$ in one of our other studies (van Teffelen et al., 2020). Fourth, to measure general psychiatric symptoms in the past 3 days, we administered the Kessler Psychological Distress Scale (K10; Kessler et al., 2002). This 10-item K10 is scored on a 5-point Likert scale ranging from 1 (*always*) to 5 (*never*; $\alpha = .95$). Finally, active alcohol and drug use were measured with the timeline follow-back questionnaire (Sobell & Sobell, 1990). All scales have demonstrated good reliability and adequate validity (Buss & Perry, 1992; Donker et al., 2010; Meesters et al., 1996; Sobell & Sobell, 1990; Verona et al., 2008).

Trait Measures. To measure different aspects of trait hostility participants completed four different measures. First, participants completed the hostility scale of the Personality Inventory for DSM-5 (PID-5 H). The 10-item PID-5 H measures overall trait hostility and was scored on a 4-point Likert scale ranging from 0 (*very false or often false*) to 3 (*very true or often true*; $\alpha = .85$). Second, we administered the AQ-H. The AQ-H is a trait variant of the AQ-HS and measures cognitive aspects of trait hostility ($\alpha = .81$). Third, participants filled in the FOA. The FOA is a trait variant of the FOAS and measures self-reported trait aggression ($\alpha = .93$). Last, we administered the voodoo doll task (VDT; DeWall et al., 2013). The VDT measures behavioral aggression. During the VDT participants are presented with an ambiguous vignette on a computer (e.g., "You are carrying a heavy load of groceries up to a check-out line at the grocery store and just as you are about to enter in line, someone cuts in front of you. You end up dropping some things on the floor"; Tremblay & Belchevski, 2004). Participants were then

[3] Study 1 was not pre-registered.

[4] We decided to deviate from pre-registration in recruiting people from the community, as patients with clinically relevant hostility levels turned out to be less agreeable in participating in our study than we hoped.

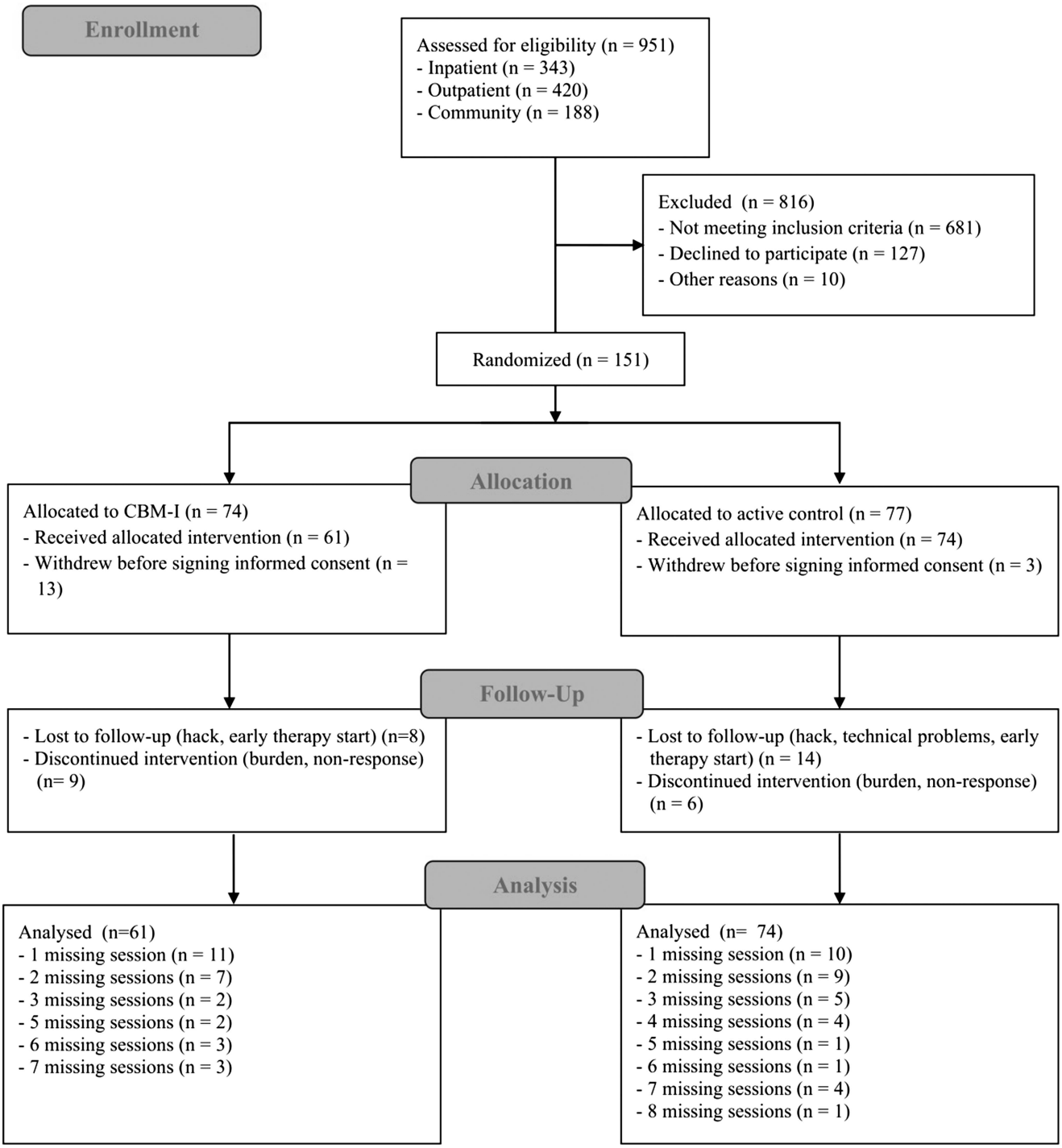
[5] People from the community were screened for eligibility by e-mail.

[6] Our CBM-I intervention took place over the course of 4 weeks. To avoid confounding effects of other therapies, people with a waiting time shorter than 4 weeks were excluded.

[7] In the pre-registration, we report mania/psychosis and therapy dropout as a measured variable. As it was not possible to screen for psychosis/mania using clinical instruments, we decided to exclude patients in case their treating clinicians reported these symptoms. Also, we stated that we would analyze drop-out ratings in participants who continued into psychotherapy after CBM-I. For the small sample that continued into psychotherapy, no drop-out was observed.

[8] We thank one of the reviewers of this paper for the valuable suggestion to include this in our manuscript.

Figure 1
Study 2—Participant Flow Diagram



Note. Unfortunately, some participants were lost to follow-up due to a successful cyberattack on Maastricht University on December 23, 2019; these participants are labeled as "hack."

allowed to insert up to 51 pins into a voodoo doll that represents the other person from the vignette. The PID-5 H, AQ-H, and the VDT have demonstrated good reliability and adequate validity (DeWall et al., 2013; Krueger et al., 2012).

Working Alliance. To measure potential carry-over effects of CBM-I on working alliance in subsequent psychotherapy we administered the Working Alliance Inventory (WAI), client version (Horvath & Greenberg, 1989), which measures the quality of the

Table 2
Study 2—Sample Characteristics

Variable	Total sample (N = 135)	Condition: CBM-I (n = 61)	Condition: AC (n = 74)	Statistical dif. χ^2/F (p)
Population sample, *n* (%)				3.14 (.370)
Inpatient	21 (16)	7 (11)	14 (19)	
Outpatient	30 (22)	14 (23)	16 (22)	
Community	84 (62)	40 (66)	44 (59)	
Primary DSM-5 classification				11.16 (.132)
Major depressive disorder	16 (12)	6 (2)	10 (14)	
Anxiety disorder	2 (1)	0 (0)	2 (3)	
Addiction	8 (6)	2 (3)	6 (8)	
Post-traumatic stress disorder	5 (4)	1 (2)	4 (5)	
Personality disorder	15 (11)	11 (18)	4 (5)	
Intermittent explosive disorder	3 (2)	1 (2)	2 (3)	
ADHD	1 (1)	0 (0)	1 (1)	
None	85 (63)	40 (66)	45 (61)	
Age in years, *M* (*SD*)	39.24 (11.80)	39.74 (12.37)	38.82 (11.37)	.45 (.656)
Gender, *n* (%)				.14 (.710)
Male	82 (61)	36 (59)	46 (62)	
Female	53 (39)	25 (41)	28 (38)	
Education, *n* (%)				.31 (.855)
Low	14 (10)	7 (11)	7 (9)	
Middle	58 (43)	27 (44)	31 (42)	
High	63 (47)	27 (44)	36 (49)	
Work situation, *n* (%)				.12 (.990)
Employed	89 (66)	41 (67)	48 (65)	
Unemployed	37 (27)	16 (26)	21 (28)	
Student	7 (5)	3 (5)	4 (5)	
Retired	2 (1)	1 (2)	1 (1)	
Medication				
AD, SSRI, *n* (%)	17 (13)	6 (10)	11 (15)	.77 (.381)
AD, SNRI, *n* (%)	4 (3)	1 (2)	3 (4)	.68 (.410)
AD, TCA, *n* (%)	5 (4)	3 (5)	2 (3)	.46 (.498)
Mood stabilizer, *n* (%)	2 (1)	1 (2)	1 (1)	.02 (.890)
Addiction, *n* (%)	1 (1)	0 (0)	1 (1)	.83 (.362)
AP, classic, *n* (%)	3 (2)	2 (3)	1 (1)	.57 (.450)
AP, atypical, *n* (%)	5 (4)	0 (0)	5 (7)	4.28 (.039)
Anxiolytic, *n* (%)	10 (7)	3 (5)	7 (9)	1.41 (.494)
Stimulant, *n* (%)	5 (4)	2 (3)	3 (4)	.06 (.812)
Aggression, *n* (%)	1 (1)	0 (0)	1 (1)	.83 (.362)
Alcohol abuse, *n* (%)				11.14 (.347)
Low	5 (4)	4 (7)	1 (1)	
Moderate	5 (4)	3 (5)	2 (3)	
High	9 (7)	4 (7)	5 (7)	
Cannabis abuse, *n* (%)				4.55 (.473)
Low	2 (1)	2 (3)	0 (0)	
Moderate	3 (2)	2 (3)	1 (1)	
High	2 (1)	1 (2)	1 (1)	

Note. CBM-I = cognitive bias modification for interpretation; AC = active control condition; AD = antidepressant medication; AP = antipsychotic medication; SSRI = selective serotonin reuptake inhibitor; SNRI = selective serotonin and noradrenaline reuptake inhibitor; TCA = tricyclic antidepressant; ADHD = attention deficit-hyperactivity disorder.

therapeutic working alliance between patient and therapist. The 36-item WAI was scored on a 5-point Likert scale ranging from 1 (*never*) to 5 (*always*). The WAI shows adequate reliability—in the present study α = .93- and adequate criterion validity (Horvath & Greenberg, 1989).

Procedure

The Ethical Review Committee Psychology and Neuroscience at Maastricht University provided ethical approval to carry out the study (ERCPN_ 170_09_11_2014_A6). This study was preregistered at https://osf.io/r46jn. The study was registered semi-retrospectively, after seven participants were recruited. Recruitment took place from September 3, 2018 until May 11, 2020. The procedure was similar to Study 1, except for four notable differences. First, the first session from this study was carried out at home instead of in our lab. Second, Study 2 was conducted as a double-blind study, instead of a single-blind study, implying that both patients and experimenters were blind for allocated condition. Third, we administered benign and hostile interpretation bias, state, and trait measures instead of solely benign and hostile interpretation bias and state affect measures. All instruments were measured

Table 3
Study 2—Descriptives

Variable	CBM-I (n = 61)		AC (n = 74)	
	Pre	Post	Pre	Post
Bias				
SIP-AEQ hostile interpretation bias	15.90 (9.17)	11.37 (8.67)	15.51 (8.15)	14.65 (8.68)
SIP-AEQ benign interpretation bias	11.56 (4.42)	14.90 (4.46)	11.37 (4.07)	11.73 (4.46)
WSAP-H hostile interpretation bias	3.43 (.81)	2.40 (1.04)	3.41 (.92)	3.15 (.88)
WSAP-H benign interpretation bias	3.81 (.99)	4.36 (.84)	3.63 (.93)	3.67 (.72)
State measures				
Hostile thoughts (AQ-HS)[a]	16.68 (8.24)	16.05 (9.01)	19.02 (8.50)	15.20 (8.48)
Anger (STAXI-2S)[a]	29.27 (10.87)	24.32 (8.82)	28.00 (10.45)	23.59 (11.67)
Aggressive behavior (FOAS)[a]	15.93 (3.81)	14.20 (3.74)	14.82 (3.43)	14.18 (4.02)
General symptoms (K10)[a,c]	34.95 (10.75)	37.56 (12.21)	34.78 (11.07)	40.67 (10.34)
Trait hostility				
Overall trait hostility (PID-5 H)	1.74 (.50)	1.52 (.55)	1.71 (.58)	1.57 (.68)
Hostile intent (AQ-H)	23.24 (6.72)	15.66 (7.66)	23.76 (6.60)	15.51 (8.19)
SR Trait aggression (FOA)[a]	71.73 (19.88)	53.73 (13.69)	66.41 (13.60)	51.06 (13.95)
Aggressive behavior (VDT)[a]	10.17 (10.93)	7.00 (8.99)	11.08 (12.14)	11.22 (13.35)
Working alliance (WAI)[b]	—	157.13 (11.10)	—	146.70 (20.34)

Note. CBM-I = cognitive bias modification for interpretation; AC = active control condition; SIP-AEQ = social information processing-attribution and emotional processing questionnaire; WSAP-H = word-sentence association paradigm hostility questionnaire; AQ-HS = aggression questionnaire-hostility subscale, state version; STAXI-2S = state-trait anger expression inventory-2; FOAS = forms of aggression questionnaire, state version; K10 = Kessler psychological distress scale-10; PID-5 H = personality inventory for DSM-5—hostility subscale; AQ-H = aggression questionnaire, hostility subscale; FOA = forms of aggression questionnaire; VDT = voodoo-doll task; WAI = working alliance inventory; SR = self-reported.
[a] Non-normally distributed. [b] FU measurement for people who engaged in therapy, n CBM-I = 8, n AC = 9. [c] Higher scores indicate less psychiatric symptoms.

pre- and post-intervention. Fourth, we monitored if participants would enter therapy after our intervention. When this occurred, we offered one additional "booster" CBM-I or AC intervention session to account for variability in time between the end of the interventions and the start of therapy. The booster sessions consisted of 33 additional scenario training and 38 additional WSAP trials. Five weeks after start of therapy, we administered the WAI once. To examine perceived usefulness, we asked participants at the end of the experiment whether they found the intervention useful (i.e., yes or no).

Statistical Analyses

Statistical analyses were similar to Study 1, except for a few notable differences. First, the hypotheses that CBM-I relates to greater increases in benign interpretation bias and greater reductions in hostile interpretation bias, state hostility, general psychiatric symptoms, and trait hostility were assessed using 11 mixed regression models. Condition, time (pre- and post-intervention), and condition (intervention and AC) by time interaction indicators were entered as predictors. Within each regression model, repeated measures were clustered in participants. AR1 was selected as covariance structure, as −2 log likelihood testing showed that it was most parsimonious. As VDT scores were highly skewed, we used Poisson regression.[9] This is in line with analytic practices in de VDT literature (DeWall et al., 2013). VDT post-score was entered as dependent variable. VDT pre-score and condition were entered as independent variables. For this specific analysis, VDT pins were imputed following the multiple imputation method. Specifically, pre-test VDT scores were used to predict post-test VDT scores in five pooled imputations.[10] Moreover, we calculated a reliable change index (RCI) for outcome variables that showed statistically significant training effects (WSAP-H and SIP-AEQ; Jacobson & Truax, 1991). Second, to examine effects of intervention on working alliance in subsequent psychotherapy one independent samples t-test on working alliance was run to test intervention differences. Analyses were conducted following the intent-to-treat principle (i.e., including participants who dropped out). Missing data were handled using maximum likelihood estimation. Multiple testing was corrected for by using the Benjamini and Hochberg (1995) procedure.[11]

Results

Means and standard deviations of study variables are presented in Table 3. Tests of baseline differences showed that variables did not differ significantly at baseline between conditions, except of atypical antipsychotic use. Overall, this indicates that random allocation was successful. Comparing our baseline values to other studies showed that hostility levels in this sample are comparable to or larger than other studies using clinical samples (Bach et al., 2016; Coccaro et al., 2017; Dillon et al., 2016; Hornsveld et al., 2009; Lievaart

[9] Pre-registration file stated that we would analyze VDT scores using mixed regression. However, VDT-scores were extremely skewed.
[10] This method involves the use of regression modeling to predict post-VDT scores based on pre-VDT scores. This prediction is made a number of times (in our case: five times, similar to common imputation practice). These five imputations are then pooled into one to reduce prediction error.
[11] Applying a multiple testing-correction was not stated in the pre-registration.

et al., 2016). In total, 14.32% of values were missing. Missed sessions per conditions are shown in Figure 1. Number of missed sessions did not differ per condition $\chi^2 = .03$, $p = .871$ and Little's MCAR test indicated that they were missing completely at random $\chi^2 = .49$, $p = .975$. At post-test, within the CBM-I and AC conditions, respectively 72.5% and 40.0% of participants indicated they found the intervention useful. Participants in the CBM-I condition perceived the intervention as more useful than participants in the AC condition ($\chi^2 = 11.80$, $p < .001$), and the indicated usefulness per condition (i.e., interaction) did not depend on hostility level ($OR = .32$, $p = .171$). To test if our results were robust to expectancy effects, analyses were run with and without covarying for perceived usefulness. As perceived usefulness did not influence the pattern of findings, results are presented without covarying for perceived usefulness. We present the results for benign and hostile interpretation, state hostility, general psychiatric symptoms, trait hostility, and working alliance below.

Benign and Hostile Interpretation Bias

To test the main hypothesis that CBM-I is related to a larger increase in benign interpretation bias and a larger decrease in hostile interpretation bias four mixed regression models were run. Fixed (i.e., reference-coded) effects of benign and hostile interpretation bias are presented in Table 4. The effects of WSAP-H benign and hostile interpretation bias are shown in and Figures 2 and 3. In Table 4, time and condition variables were reference coded using the AC condition as reference. Hence, fixed effects presented in Table 4 are estimated using AC as reference category. As the interaction effects in Table 4 show, findings showed that CBM-I led to a greater increase in SIP-AEQ ($d = 0.29$, 95% CI = 0.12 to 0.46) and WSAP-H ($d = 0.26$, 95% CI = 0.09 to 0.43) benign interpretation bias from pre- to post-intervention compared to AC. Results also showed that CBM-I led to a greater decrease in WSAP-H ($d = -0.33$, 95% CI = −0.83 to −0.17), but not in SIP-AEQ hostile interpretation bias from pre- to post-intervention compared to AC. On the SIP-AEQ, 23.0% of participants in the CBM-I group and 6.8% of participants in the AC group showed significant reliable change ($\chi^2 = 7.25$, $p = .007$). On the WSAP-H, 57.4% of participants in the CBM-I group and 32.4% of participants in the AC group showed significant reliable change ($\chi^2 = 8.46$, $p = .004$).

State Hostility and General Psychiatric Symptom Measure

To test the hypothesis that CBM-I would lead to larger reductions in state hostility and general psychiatric symptoms than AC, four separate mixed regression models were run with cognitive aspects of hostility (AQ-H), state anger (STAXI-2S), self-reported aggression (FOAS), and general psychiatric symptoms (K10) as dependent variables and time, condition, and condition by time as independent variables. Findings are shown in Table 4. Results on self-reported state aspects of hostility (AQ-H, STAXI-2S, and FOAS) and general psychiatric symptoms showed that none of the interaction effects of condition with time were significant ($p > .088$).

Trait Hostility Measures

To test the prediction that CBM-I was related to larger reductions in hostility traits than AC, four separate mixed regression models were run with overall hostility (PID5-H), cognitive aspects of hostility (AQ-H), and self-reported aggression traits (FOA) as dependent variables and time, condition, and condition by time as independent variables. Results are shown in Table 4. For the self-reported hostility measures (PID-5 H, AQ-H, and FOA), findings showed no significant interaction effects ($p > .244$). Next, a Poisson regression model was run with behavioral aggression (VDT) at post-intervention as dependent variable and condition and behavioral aggression at baseline as independent variables. Results showed that CBM-I was associated with greater reductions in behavioral aggression from pre- to post-intervention ($B = -.28$, $SE = .08$, $d = -0.29$, 95% CI = −.47 to −.10) compared to AC (see Figure 4). On the VDT, 9.7% of participants in the CBM-I group and 13.8% of participants in the AC group showed significant reliable change ($\chi^2 = .57$, $p = .452$).

Explorative Analyses

Given that men express aggression more physically than women (Björkqvist et al., 1992), we explored CBM-I effects on behavioral aggression for men. Results showed that the effect of CBM-I increased ($d = -0.41$, 95% CI = −1.10 to −0.20) compared to AC.

Carry-Over Effects on Working Alliance in Subsequent Psychotherapy

To test the hypothesis that CBM-I relates to beneficial carry-over effects on working alliance in subsequent psychotherapy we explored the data of $n = 17$ ($n = 8$ for CBM-I and $n = 9$ for AC) participants who engaged in psychotherapy after the experiment. First, an independent samples t-test was run on working alliance as dependent variable. Results demonstrated that the two conditions did not differ significantly in terms of working alliance in psychotherapy, $F(1, 96.09) = .33$, $p = .671$.

Study 2—Discussion

In Study 2 we tested the efficacy of CBM-I intervention versus AC in a larger double-blind, sham-controlled clinical study in people with clinical levels of hostility where the additional impact on hostility outcomes and general psychiatric symptoms was assessed. Study 2 showed that compared to AC, CBM-I increased benign interpretation bias, and partially reduced hostile interpretation bias and aggressive behavior (but not self-reported aggression). Also, findings showed no condition differences on self-reported state and trait hostility, and general psychiatric symptom measures. Moreover, explorative findings showed no difference between conditions in working alliance for people who went into psychotherapy after our interventions.

General Discussion

The present work investigated the effects of a CBM-I intervention for hostility. We first tested its feasibility in a single-blind randomized sham-controlled feasibility study using a mixed clinical-community male sample ($N = 29$, Study 1) and then tested its efficacy in a double-blind, randomized sham-controlled clinical study in a mix-gender sample of people with clinical levels of

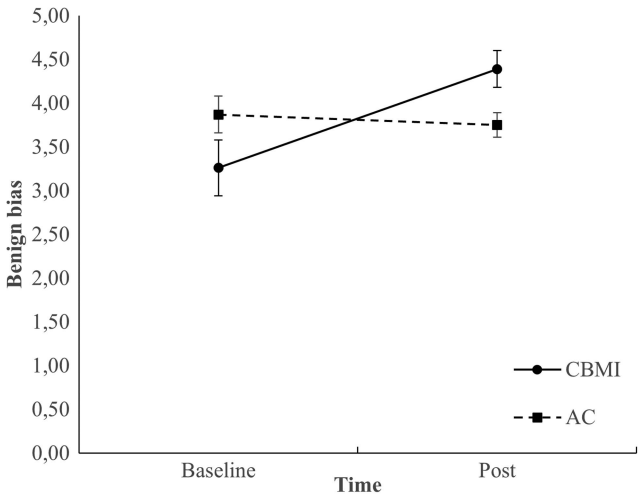
Table 4
Study 2—Fixed Effects of Mixed Regression on Study Variables

Variable	*b*	*SE*	*t*	*p*
Bias				
SIP-AEQ Benign interpretation bias				
Intercept	10.64	.91	11.68	<.001***
Time	−.58	.57	1.02	.370
Condition	−2.87	1.37	−2.09	.076
Condition by time	2.87	.86	3.34	<.001***
SIP-AEQ Hostile interpretation bias				
Intercept	17.47	1.65	10.60	<.001***
Time	−1.56	.98	−1.61	.163
Condition	2.99	2.49	1.20	.290
Condition by time	−2.94	1.48	−1.99	.088
WSAP-H Benign interpretation bias				
Intercept	3.87	.21	18.09	<.001***
Time	−.12	.14	−.88	.432
Condition	−.61	.32	−1.88	.101
Condition by time	.64	.21	3.05	.007**
WSAP-H Hostile interpretation bias				
Intercept	3.73	.20	19.07	<.001***
Time	−.31	.12	−2.53	.027*
Condition	.77	.29	2.60	.022*
Condition by time	−.73	.19	−3.95	<.001***
State measures				
AQ-HS				
Intercept	23.14	1.61	14.34	<.001***
Time	−3.90	.99	−3.95	<.001***
Condition	−4.28	2.41	−1.77	.118
Condition by time	2.85	1.46	1.95	.091
STAXI-2S				
Intercept	35.07	2.31	15.19	<.001***
Time	−5.26	1.47	−3.59	.002**
Condition	.37	3.45	.11	.935
Condition by time	−.08	2.18	−.04	.970
FOAS				
Intercept	16.77	.80	20.91	<.001***
Time	−.97	.49	−2.00	.088
Condition	1.70	1.20	−2.00	.226
Condition by time	−.97	.72	−1.34	.183
K10				
Intercept	27.28	1.89	14.42	<.001***
Time	6.20	1.07	5.81	<.001***
Condition	2.77	2.83	.98	.381
Condition by time	−3.16	1.58	−1.99	.088
Trait hostility				
PID-5 H				
Intercept	1.83	.11	16.77	<.001***
Time	−.08	.06	−1.31	.247
Condition	.14	.16	.83	.451
Condition by time	−.11	.09	−1.17	.297
AQ-H				
Intercept	32.81	1.37	23.96	<.001***
Time	−8.37	.79	−10.56	<.001***
Condition	−.30	2.05	−.15	.925
Condition by time	.56	1.20	.47	.688
FOA				
Intercept	84.51	3.24	26.05	<.001***
Time	−14.94	1.90	−7.85	<.001***
Condition	8.71	4.88	1.79	.118
Condition by time	−3.90	2.89	−1.35	.244

Note. In all models the active control condition was chosen as reference category. SIP-AEQ = social information processing-attribution and emotional processing questionnaire; WSAP-H = word-sentence association paradigm hostility questionnaire; AQ-HS = aggression questionnaire-hostility subscale, state version; STAXI-2S = state-trait anger expression inventory-2; FOAS = forms of aggression questionnaire, state version; K10 = Kessler psychological distress scale-10; PID-5 H = personality inventory for DSM-5—hostility subscale; AQ-H = aggression questionnaire, hostility subscale; FOA = forms of aggression questionnaire.
* $p < .050$. ** $p < .010$. *** $p < .001$.

Figure 2

Study 2—Estimates of WSAP-H Benign Interpretation Bias Over Time Per Condition

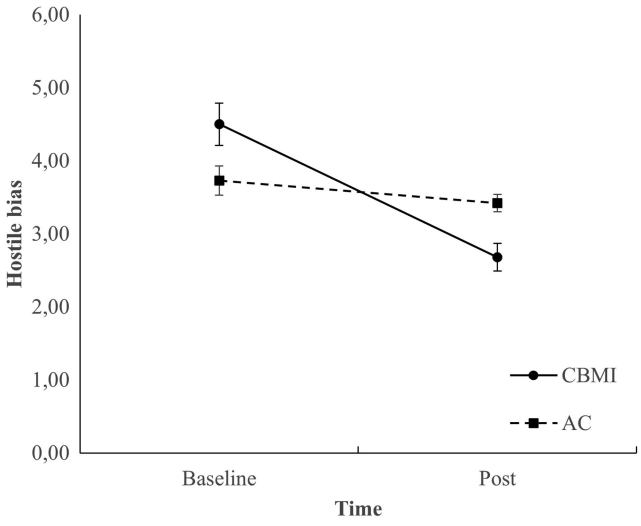


Note. Error bars represent standard errors of the estimates. We chose one outcome (WSAP) to depict as this strongly resembles the SIP-AEQ pattern. $N = 135$.

hostility ($N = 135$, Study 2). Overall, we found that eight sessions of CBM-I across 4 weeks increased benign interpretation bias. Study 1 indicated moderate to good feasibility but showed no significant effects on hostile interpretation bias. In Study 2, CBM-I increased benign interpretation bias with small effect sizes. The hypothesis that CBM-I reduces hostile interpretation bias and aggression found only partial support. However, we observed no differential changes in self-reported hostility measures and general psychiatric symptoms. Furthermore, we did not observe a differential impact of CBM-I on working alliance in subsequent psychotherapy; however, this analysis was very underpowered.

Figure 3

Study 2—Estimates of WSAP-H Hostile Interpretation Bias Over Time Per Condition



Note. Error bars represent standard errors of the estimates. $N = 135$.

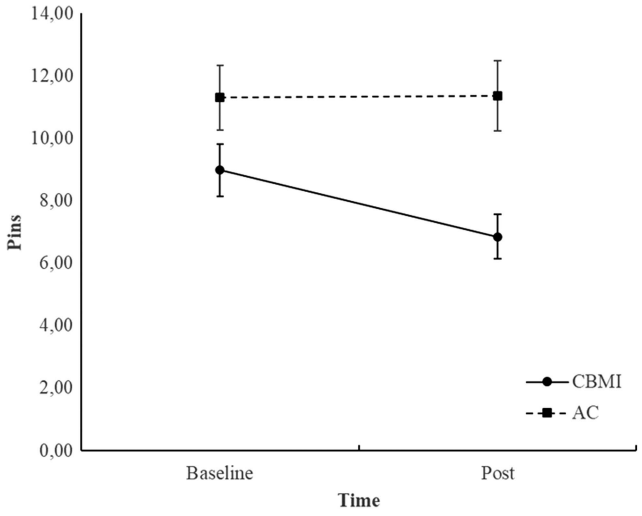
The strength of therapy effects on benign and hostile interpretation bias in the present study (i.e., $d = 0.29$ and $d = 0.26$ for benign interpretation bias, and $d = -0.33$ for hostile interpretation bias), are smaller than those in a number of previous studies mentioned in the introduction. To compare, the study that compared one CBM-I session to a sham condition in students and found medium effect sizes ($d = 0.44$ for benign interpretation bias and $d = 0.66$ for hostile interpretation bias; Hawkins & Cougle, 2013). The two studies that compared an eight-session CBM-I training to a sham condition and demonstrated medium to large effect sizes ($d = 1.17$ for benign interpretation bias and $d = 0.65$ for hostile interpretation bias) in non-treatment seeking people with alcohol use disorder (Cougle et al., 2017) and major depressive disorder ($d = 1.06$ for benign interpretation bias and $d = 0.64$ for hostile interpretation bias; Smith et al., 2018). The study that compared an eight CBM-I sessions to a waiting list condition in aggressive boys found large effect sizes ($d = 1.40$ for benign interpretation bias and $d = 2.21$ for hostile interpretation bias; Vassilopoulos et al., 2015). Last, the study that compared one session of CBM-I benign training to a negative training condition showed a large effect size (approximate $d = 0.85$ for benign interpretation bias; AlMoghrabi et al., 2018). However, the effect size on aggression in the current study does converge with findings of the recent meta-analysis across 85 studies showing that CBM-I for anxiety ($SMD = -0.30$) and depression ($SMD = -0.26$) demonstrates small effects on symptoms compared to a sham condition (Fodor et al., 2020). Importantly, this meta-analysis showed that studies with lower methodological quality and therefore higher risk of bias generally found higher effect sizes. Accordingly, an explanation for the smaller effect size in our study may be that it fulfills criteria for high methodological quality and low risk of bias (e.g., intention to treat analysis, random sequence generation, allocation concealment, blinding).

Contrary to predictions, we observed no effect of CBM-I on hostile interpretation bias in Study 1 and only partial efficacy of CBM-I on hostile interpretation bias in Study 2. One explanation for this finding is the smaller sample size in Study 1, which had lower power. Another explanation is that the observed effects in CBM-I studies depend on the specificity of the included measurement instruments. In essence, CBM-I intends to train more *benign* interpretations. Changes in interpretation bias are then more likely to be observed in instruments that measure *benign*, as opposed to *non-benign* interpretation tendencies. For example, one study in patients with major depressive disorder showed that two sessions of CBM-I across 7 days increased positive interpretation bias, but had no significant effect on depressive interpretation bias compared to healthy control participants (Joormann et al., 2015).

Next to an observed efficacy on interpretation bias, Study 2 showed only partial efficacy of CBM-I on aggression. Specifically, we observed a small effect of CBM-I on aggressive *behavior*, but not on self-reported aggression. One explanation for the observed pattern of findings is that we have found a false-positive effect of behavioral aggression, and that CBM-I has no beneficial effect on aggression. That being said, human aggression is difficult to measure, and it is not uncommon for self-reported and behavioral aggression to correlate differentially with outcome measures (Brugman et al., 2015; Giancola & Parrott, 2008; Hyatt et al., 2019). Alternatively, the finding that CBM-I reduces behavioral

Figure 4

Study 2—VDT Pins Over Time Per Condition



Note. Error bars represent standard errors of the estimates. VDT = Voodoo Doll Task. $N = 135$.

hostility may suggest an information processing pathway towards aggressive behavior that operates distinctly from non-behavioral aspects of hostility. This is in line with the Social Information Processing model of aggression (Crick & Dodge, 1994), which implies that the hostile interpretation of external (i.e., situational) and internal (e.g., emotional) cues results in a narrowing of potential behavioral response patterns. This, in turn, increases the likelihood of aggressive response patterns.

The explorative finding that working alliance levels in subsequent psychotherapy did not differ significantly between conditions could indicate a true negative finding. That is, the effects of CBM-I on bias levels may be too small to instigate a carryover effect. Notwithstanding, statistical power in the present explorative sample was simply too small to conclude that there is no true effect of CBM-I on the perceived quality of the working alliance in subsequent psychotherapy. We cannot rule out that CBM-I may have augmenting effects when provided prior to treatment, but this issue is largely neglected in the field and requires further experimental evidence.

Several important limitations impact the present findings. First, the samples included participants with clinical levels of hostility that were both treatment- and non-treatment-seeking people. Although most people were screened in a treatment-seeking population (80%), treatment-seeking people made up (only) 38% of the final sample. It could be argued that people with clinical levels of hostility are more likely to decline participation, but Figure 1 shows that our sample composition is more likely due to treatment-seeking people that did not meet our inclusion criteria. However, this still implies that our results may not generalize to a treatment-seeking sample that for instance shows increased numbers of experienced psychosocial stressors, comorbidity, tendencies for interpersonal conflict, and premature treatment discontinuation. As a related issue, the present sample included more men than women. Given that men express aggression more physically than women (Björkqvist et al., 1992; Genovese et al., 2017) this could have amplified the strength of the present findings. When we explored CBM-I effects on behavioral aggression for men only, its effect slightly increased from $d = 0.28$ using the complete sample to $d = 0.41$. A second limitation of the present study is that it did not include a follow-up measurement. We therefore do not know whether the results sustain over time. Third, there is currently no consensus in the CBM-I literature in terms of dose-response effects. We opted for eight 20-min sessions, but we urge future studies to investigate the optimal dose-response effect. A fourth drawback of the study was that we omitted to define interpretation bias as a main outcome prior to the study, while we did base the a priori power analysis on interpretation bias only. Fifth, the present study contained an active and a control condition. Perhaps, the observed findings were impacted by the fact that people who score high on hostility do not like to randomly allocated to experimental groups. However, participants were explicitly told they were going to be randomly allocated to an intervention study for hostility with active or control condition. We also asked the participants whether or not they thought the interventions were beneficial for them. Within the CBM-I and AC conditions, respectively 72.5% and 40.0% of participants indicated they found the intervention useful. However, outcomes were not affected when we covaried for perceived usefulness. Last, we originally intended to exclude people who actively used alcohol or drugs. After additional scrutiny of the literature, however, we could not find convincing evidence that supported this criterion. On the contrary, literature showed that bias modification studies are conducted and shown to be efficacious in samples that are on active alcohol use (Wiers et al., 2015). The latter study showed for example that the alcohol approach bias significantly reduced, but non-differentially from active control training. We recommend future studies on CBM-I to further disentangle the influence of alcohol and substance consumption on CBM-I efficacy.

The present work holds several clinical implications. First, the finding that CBM-I for hostility increased benign interpretation bias and partly reduced hostile interpretation bias and aggression implies that people with clinical hostility levels experience small but significant improvements after the repeated stimulation of benign interpretations in random ambiguous scenarios. This shows that CBM-I holds promise as a prevention or intervention strategy for hostility at relatively low cost. In addition, explorative analysis of CBM-I effects on (physical) aggressive behavior suggests that the effect is slightly stronger for men. Evidently, this finding suggestion requires replication. Our findings also suggest that CBM-I could be implemented in both treatment- and non-treatment-seeking settings. However, research is still in its' early stages as a number of important questions remain unanswered at this point: Does efficacy sustain over time? What is the optimal dose-response effect? Can CBM-I serve as an add-on to standard treatment, for example when people are on waiting list? And, is the intervention effective in everyday clinical practice? The questions require further research prior to further implementation.

Overall, we found modest support for CBM-I as an intervention for hostility, with some evidence of its efficacy for hostile interpretation bias and aggression. We discuss study limitations as well as directions for future research.

The data in this study has not been published before and is currently not in press or under review.

References

AlMoghrabi, N., Huijding, J., & Franken, I. H. A. (2018). The effects of a novel hostile interpretation bias modification paradigm on hostile interpretations, mood, and aggressive behavior. *Journal of Behavior Therapy and Experimental Psychiatry*, *58*, 36–42. https://doi.org/10.1016/j.jbtep.2017.08.003

Ammerman, B. A., Kleiman, E. M., Uyeji, L. L., Knorr, A. C., & McCloskey, M. S. (2015). Suicidal and violent behavior: The role of anger, emotion dysregulation, and impulsivity. *Personality and Individual Differences*, *79*, 57–62. https://doi.org/10.1016/j.paid.2015.01.044

Arntz, A., Stupar-Rutenfrans, S., Bloo, J., van Dyck, R., & Spinhoven, P. (2015). Prediction of treatment discontinuation and recovery from Borderline Personality Disorder: Results from an RCT comparing Schema Therapy and Transference Focused Psychotherapy. *Behaviour Research and Therapy*, *74*, 60–71. https://doi.org/10.1016/j.brat.2015.09.002

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Bach, B., Maples-Keller, J. L., Bo, S., & Simonsen, E. (2016). The alternative DSM–5 personality disorder traits criterion: A comparative examination of three self-report forms in a Danish population. *Personality Disorders: Theory, Research, and Treatment*, *7*(2), 124–135. https://doi.org/10.1037/per0000162

Beard, C., & Amir, N. (2008). A multi-session interpretation modification program: Changes in interpretation and social anxiety symptoms. *Behaviour Research and Therapy*, *46*(10), 1135–1141. https://doi.org/10.1016/j.brat.2008.05.012

Beard, C., Rifkin, L. S., Silverman, A. L., & Björgvinsson, T. (2019). Translating CBM-I into real-world settings: Augmenting a CBT-based psychiatric hospital program. *Behavior Therapy*, *50*(3), 515–530. https://doi.org/10.1016/j.beth.2018.09.002

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series A, (Statistics in Society)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Björkqvist, K., Lagerspetz, K. M., & Kaukiainen, A. (1992). Do girls manipulate and boys fight? Developmental trends in regard to direct and indirect aggression. *Aggressive Behavior*, *18*(2), 117–127. https://doi.org/10.1002/1098-2337(1992)18:2<117::AID-AB2480180205>3.0.CO;2-3

Bowler, J. O., Mackintosh, B., Dunn, B. D., Mathews, A., Dalgleish, T., & Hoppitt, L. (2012). A comparison of cognitive bias modification for interpretation and computerized cognitive behavior therapy: Effects on anxiety, depression, attentional control, and interpretive bias. *Journal of Consulting and Clinical Psychology*, *80*(6), 1021–1033. https://doi.org/10.1037/a0029932

Brugman, S., Lobbestael, J., Arntz, A., Cima, M., Schuhmann, T., Dambacher, F., & Sack, A. T. (2015). Identifying cognitive predictors of reactive and proactive aggression. *Aggressive Behavior*, *41*(1), 51–64. https://doi.org/10.1002/ab.21573

Buss, A. H., & Perry, M. (1992). The aggression questionnaire. *Journal of Personality and Social Psychology*, *63*(3), 452–459. https://doi.org/10.1037/0022-3514.63.3.452

Butler, E., Mobini, S., Rapee, R. M., Mackintosh, B., & Reynolds, S. A. (2015). Enhanced effects of combined cognitive bias modification and computerised cognitive behaviour therapy on social anxiety. *Cogent Psychology*, *2*(1). Advance online publication. https://doi.org/10.1080/23311908.2015.1011905

Cassiello-Robbins, C., & Barlow, D. H. (2016). Anger: The unrecognized emotion in emotional disorders. *Clinical Psychology: Science and Practice*, *23*(1), 66–85. https://doi.org/10.1111/cpsp.12139

Cassiello-Robbins, C., Conklin, L. R., Anakwenze, U., Gorman, J. M., Woods, S. W., Shear, M. K., & Barlow, D. H. (2015). The effects of aggression on symptom severity and treatment response in a trial of cognitive behavioral therapy for panic disorder. *Comprehensive Psychiatry*, *60*, 1–8. https://doi.org/10.1016/j.comppsych.2015.04.012

Coccaro, E. F., Fanning, J., & Lee, R. (2017). Development of a social emotional information processing assessment for adults (SEIP-Q). *Aggressive Behavior*, *43*(1), 47–59. https://doi.org/10.1002/ab.21661

Coccaro, E. F., Noblett, K. L., & McCloskey, M. S. (2009). Attributional and emotional responses to socially ambiguous cues: Validation of a new assessment of social/emotional information processing in healthy adults and impulsive aggressive patients. *Journal of Psychiatric Research*, *43*(10), 915–925. https://doi.org/10.1016/j.jpsychires.2009.01.012

Cougle, J. R., Summers, B. J., Allan, N. P., Dillon, K. H., Smith, H. L., Okey, S. A., & Harvey, A. M. (2017). Hostile interpretation training for individuals with alcohol use disorder and elevated trait anger: A controlled trial of a web-based intervention. *Behaviour Research and Therapy*, *99*, 57–66. https://doi.org/10.1016/j.brat.2017.09.004

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. https://doi.org/10.1037/0033-2909.115.1.74

Del Vecchio, T., & O'Leary, K. D. (2004). Effectiveness of anger treatments for specific anger problems: A meta-analytic review. *Clinical Psychology Review*, *24*(1), 15–34. https://doi.org/10.1016/j.cpr.2003.09.006

DeWall, C. N., Finkel, E. J., Lambert, N. M., Slotter, E. B., Bodenhausen, G. V., Pond, R. S., Jr., Renzetti, C. M., & Fincham, F. D. (2013). The voodoo doll task: Introducing and validating a novel method for studying aggressive inclinations. *Aggressive Behavior*, *39*(6), 419–439. https://doi.org/10.1002/ab.21496

Dillon, K. H., Allan, N. P., Cougle, J. R., & Fincham, F. D. (2016). Measuring Hostile Interpretation Bias:The WSAP-Hostility Scale. *Assessment*, *23*(6), 707–719. https://doi.org/10.1177/1073191115599052

Donker, T., Comijs, H., Cuijpers, P., Terluin, B., Nolen, W., Zitman, F., & Penninx, B. (2010). The validity of the Dutch K10 and extended K10 screening scales for depressive and anxiety disorders. *Psychiatry Research*, *176*(1), 45–50. https://doi.org/10.1016/j.psychres.2009.01.012

Fodor, L. A., Georgescu, R., Cuijpers, P., Szamoskozi, Ş., David, D., Furukawa, T. A., & Cristea, I. A. (2020). Efficacy of cognitive bias modification interventions in anxiety and depressive disorders: A systematic review and network meta-analysis. *The Lancet Psychiatry*, *7*(6), 506–514. https://doi.org/10.1016/S2215-0366(20)30130-9

Genovese, T., Dalrymple, K., Chelminski, I., & Zimmerman, M. (2017). Subjective anger and overt aggression in psychiatric outpatients. *Comprehensive Psychiatry*, *73*, 23–30. https://doi.org/10.1016/j.comppsych.2016.10.008

Giancola, P. R., & Parrott, D. J. (2008). Further evidence for the validity of the Taylor aggression paradigm. *Aggressive Behavior*, *34*(2), 214–229. https://doi.org/10.1002/ab.20235

Gülüm, İ. V., Soygüt, G., & Safran, J. D. (2018). A comparison of pre-dropout and temporary rupture sessions in psychotherapy. *Psychotherapy Research*, *28*(5), 685–707. https://doi.org/10.1080/10503307.2016.1246765

Hallion, L. S., & Ruscio, A. M. (2011). A meta-analysis of the effect of cognitive bias modification on anxiety and depression. *Psychological Bulletin*, *137*(6), 940–958. https://doi.org/10.1037/a0024355

Hawkins, K. A., & Cougle, J. R. (2013). Effects of interpretation training on hostile attribution bias and reactivity to interpersonal insult. *Behavior Therapy*, *44*(3), 479–488. https://doi.org/10.1016/j.beth.2013.04.005

Henrichs, J., Bogaerts, S., Sijtsema, J., & Klerx-van Mierlo, F. (2015). Intimate partner violence perpetrators in a forensic psychiatric outpatient setting: Criminal history, psychopathology, and victimization. *Journal of Interpersonal Violence*. Advance online publication. https://doi.org/10.1177/0886260514552272

Hirsh, J. B., Quilty, L. C., Bagby, R. M., & McMain, S. F. (2012). The relationship between agreeableness and the development of the working

alliance in patients with borderline personality disorder. *Journal of Personality Disorders*, *26*(4), 616–627. https://doi.org/10.1521/pedi.2012.26.4.616

Hofmann, S. G., Asnaani, A., Vonk, I. J. J., Sawyer, A. T., & Fang, A. (2012). The efficacy of cognitive behavioral therapy: A review of meta-analyses. *Cognitive Therapy and Research*, *36*(5), 427–440. https://doi.org/10.1007/s10608-012-9476-1

Holmes, E. A., Mathews, A., Dalgleish, T., & Mackintosh, B. (2006). Positive Interpretation Training: Effects of Mental Imagery Versus Verbal Training on Positive Mood. *Behavior Therapy*, *37*(3), 237–247. https://doi.org/10.1016/j.beth.2006.02.002

Hornsveld, R. H. J., Muris, P., Kraaimaat, F. W., & Meesters, C. (2009). Psychometric properties of the Aggression Questionnaire in Dutch violent forensic psychiatric patients and secondary vocational students. *Assessment*, *16*(2), 181–192. https://doi.org/10.1177/1073191108325894

Horvath, A. O., & Greenberg, L. S. (1989). Development and validation of the Working Alliance Inventory. *Journal of Counseling Psychology*, *36*(2), 223–233. https://doi.org/10.1037/0022-0167.36.2.223

Hyatt, C. S., Chester, D. S., Zeichner, A., & Miller, J. D. (2019). Analytic flexibility in laboratory aggression paradigms: Relations with personality traits vary (slightly) by operationalization of aggression. *Aggressive Behavior*, *45*(4), 377–388. https://doi.org/10.1002/ab.21830

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, *59*(1), 12–19. https://doi.org/10.1037/0022-006X.59.1.12

Joormann, J., Waugh, C. E., & Gotlib, I. H. (2015). Cognitive bias modification for interpretation in major depression: Effects on memory and stress reactivity. *Clinical Psychological Science*, *3*(1), 126–139. https://doi.org/10.1177/2167702614560748

Kessler, R. C., Andrews, G., Colpe, L. J., Hiripi, E., Mroczek, D. K., Normand, S. L. T., Walters, E. E., & Zaslavsky, A. M. (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological Medicine*, *32*(6), 959–976. https://doi.org/10.1017/S0033291702006074

Krueger, R. F., Derringer, J., Markon, K. E., Watson, D., & Skodol, A. E. (2012). Initial construction of a maladaptive personality trait model and inventory for DSM-5. *Psychological Medicine*, *42*(9), 1879–1890. https://doi.org/10.1017/S0033291711002674

Lievaart, M., Franken, I. H. A., & Hovens, J. E. (2016). Anger assessment in clinical and nonclinical populations: Further validation of the State–Trait Anger Expression Inventory-2. *Journal of Clinical Psychology*, *72*(3), 263–278. https://doi.org/10.1002/jclp.22253

Mathews, A., & Mackintosh, B. (2000). Induced emotional interpretation bias and anxiety. *Journal of Abnormal Psychology*, *109*(4), 602–615. https://doi.org/10.1037/0021-843X.109.4.602

McNair, D. M., Lorr, M., & Droppleman, L. F. (1992). *Manual for the profile of mood states*. Educational and Industrial Testing Service.

Meesters, C., Muris, P., Bosma, H., & Schouten, E. (1996). Psychometric evaluation of the Dutch version of the Aggression Questionnaire. *Behaviour Research and Therapy*, *34*(10), 839–843. https://doi.org/10.1016/0005-7967(96)00065-4

Putt, C. A., Dowd, E. T., & McCormick, R. A. (2001). Impact of pre-existing levels of hostility and aggression on substance abuse treatment outcome. *Counselling Psychology Quarterly*, *14*(2), 139–147. https://doi.org/10.1080/09515070110058576

Smith, H. L., Dillon, K. H., & Cougle, J. R. (2018). Modification of Hostile Interpretation Bias in Depression: A Randomized Controlled Trial. *Behavior Therapy*, *49*(2), 198–211. https://doi.org/10.1016/j.beth.2017.08.001

Sobell, L. C., & Sobell, M. B. (1990). Self-report issues in alcohol abuse: State of the art and future directions. *Behavioral Assessment*, *12*(1), 77–90.

Spielberger, C. D. (1999). *State-Trait Anger Expression Inventory-2*. Psychological Assessment Resources.

Tremblay, P. F., & Belchevski, M. (2004). Did the Instigator Intend to Provoke? A Key Moderator in the Relation Between Trait Aggression and Aggressive Behavior. *Aggressive Behavior*, *30*(5), 409–424. https://doi.org/10.1002/ab.20027

Van der Heijden, P., Ingenhoven, T., Berghuis, H., & Rossi, G. (2014). *Nederlandstalige bewerking van The Personality Inventory for DSM-5® (PID-5)—Adult, 2011*. Uitgeverij Boom.

van Teffelen, M. W., Lobbestael, J., Voncken, M. J., & Peeters, F. (2020). Uncovering the hierarchical structure of self-reported hostility. *PLOS ONE*, *15*(9), Article e0239631. https://doi.org/10.1371/journal.pone.0239631

Vassilopoulos, S. P., Brouzos, A., & Andreou, E. (2015). A multi-session attribution modification program for children with aggressive behaviour: Changes in attributions, emotional reaction estimates, and self-reported aggression. *Behavioural and Cognitive Psychotherapy*, *43*(05), 1–11. https://doi.org/10.1017/S1352465814000149

Verona, E., Sadeh, N., Case, S. M., Reed, A., II, & Bhattacharjee, A. (2008). Self-reported use of different forms of aggression in late adolescence and emerging adulthood. *Assessment*, *15*(4), 493–510. https://doi.org/10.1177/1073191108318250

von der Lippe, A. L., Monsen, J. T., Rønnestad, M. H., & Eilertsen, D. E. (2008). Treatment failure in psychotherapy: The pull of hostility. *Psychotherapy Research*, *18*(4), 420–432. https://doi.org/10.1080/10503300701810793

Wicherts, J. M., & Vorst, H. C. M. (2004). Modelpassing van de Verkorte Profile of Mood States en meetinvariantie over mannen en vrouwen [journal article]. *Nederlands Tijdschrift voor de Psychologie en haar Grensgebieden*, *59*(1), 11–20. https://doi.org/10.1007/BF03062320

Wiers, R. W., Houben, K., Fadardi, J. S., van Beek, P., Rhemtulla, M., & Cox, W. M. (2015). Alcohol cognitive bias modification training for problem drinkers over the web. *Addictive Behaviors*, *40*, 21–26. https://doi.org/10.1016/j.addbeh.2014.08.010

Received August 20, 2020
Revision received March 10, 2021
Accepted March 15, 2021 ■