APPLIED COGNITIVE PSYCHOLOGY
Appl. Cognit. Psychol. **24**: 326–337 (2010)
Published online in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/acp.1679

Does Modified Interpretation Bias Influence Automatic Avoidance Behaviour?

WOLF-GERO LANGE[1]*, ELSKE SALEMINK[2†], INE WINDEY[3], GER P. J. KEIJSERS[1], JULIE KRANS[1], ENI S. BECKER[1] and MIKE RINCK[1]

[1]*Behavioural Science Institute, Radboud University, Nijmegen, The Netherlands*
[2]*University of Utrecht, The Netherlands*
[3]*University of Leuven, Belgium*

SUMMARY

Cognitive bias modification (CBM) studies suggest a causal role of interpretation biases in the aetiology and maintenance of Social Anxiety Disorder. However, it is unknown if the effects of induced biases transfer to behaviour. In two analogue studies, behavioural changes in response to aversive and positive stimuli were measured after the induction of positive and negative interpretation biases in 'averagely anxious' participants. Responses to emotional multi-facial displays ('crowds') were measured using an indirect Approach–Avoidance Task (AAT). The crowds comprised different ratios of either neutral and angry faces or happy and angry faces. In Experiment 1, negatively trained participants (NETs) showed a faster avoidance response for the neutral–angry crowds when the number of angry pictures in the crowd increased. This response pattern resembles the one previously found in socially anxious individuals. Experiment 2 replicated the effect of the cognitive bias manipulation on conceptually comparable material, but did not show transfer to the behavioural task. These studies add to the body of knowledge regarding successful modification of interpretive bias and generalizability to a behavioural task. Copyright © 2010 John Wiley & Sons, Ltd.

A large body of evidence supports the notion that biased information processing is crucial for the aetiology and maintenance of psychiatric disorders (e.g. Clark & Wells, 1995). Negative biases in the interpretation of social information is one of the most prominent dysfunctional cognitive processes in social phobia (Foa, Franklin, & Kozak, 2001). When participants are, for instance, asked to interpret outcomes of ambiguous social situations, social anxiety seems to be associated with more negative interpretations (Huppert, Pasupuleti, Foa, & Mathews, 2007). Moreover, Voncken, Bögels, and de Vries (2003) found that socially phobics tended to interpret *all* kinds of social scenarios negatively, not just ambiguous ones.

Contrary to the straightforward findings with text materials, there is less persuasive evidence when participants have to evaluate facial expressions (e.g. Merckelbach, Van

*Correspondence to: Wolf-Gero Lange, Department of Clinical Psychology, Radboud University of Nijmegen, PO Box 9104, 6500 HE Nijmegen, The Netherlands. E-mail: g.lange@psych.ru.nl
†Now at the University of Amsterdam, The Netherlands.

Hout, Van den Hout, & Mersch, 1989; but also Philippot & Douilliez, 2005). This is rather surprising when considering that facial expressions are thought to have strong communicative (thus social) value, to leave room for interpretation with respect to the reason, intention and target of the expression, and to be evolutionarily relevant (Vuilleumier, 2002). Recently, however, there is evidence that negative evaluations of facial expressions are reflected in reflexive approach and avoidance tendencies, but not necessarily in controlled direct evaluations (Heuer, Rinck, & Becker, 2007; Lange, Keijsers, Becker, & Rinck, 2008). In both studies, participants were instructed to pull emotional faces towards themselves (approach) or push them away (avoid) by means of a joystick. Even though the depicted emotions were task-irrelevant in both studies, socially anxious participants generally showed speeded avoidance reactions to angry and happy, but not to neutral faces. When asked to rate the same faces directly in terms of friendliness, the groups did not differ.

Just recently, researchers have started to investigate the causal underpinnings of the relation between anxiety and cognitive bias. As an analogy, techniques have been developed that allow for the induction of interpretation biases or attentional biases in nonanxious participants (cognitive bias modification (CBM)). Important issues regarding CBM are (1) whether an induced bias only transfers to conceptually similar new materials, or whether it also generalizes to other domains, and (2) whether an induced bias is reflected in measures of anxiety or anxiety-related behaviour. Recent evidence indeed suggests that induced interpretation biases (CBM-I) influence subsequently reported anxiety mostly in response to stress, as predicted by the valence of the training (e.g. Mackintosh, Mathews, Yiend, Ridgeway, & Cook, 2006). Specifically, Beard and Amir (2008) reported that socially anxious individuals showed a significant drop in anxiety symptoms after eight sessions of CBM-I. Yet, to our knowledge, only one study has attempted to change anxiety-related behaviour (approaching a living spider) by CBM-I, and the results were inconclusive (Teachman & Addison, 2008).

Since Lange et al. (2008) showed that facial crowds (i.e. a matrix of faces presented on a computer monitor) trigger reflexive avoidance responses in socially anxious individuals, we hypothesized that the same should be true for people who are trained to interpret social situations negatively. Specifically, in socially anxious individuals these response tendencies are known to become increasingly avoidant with an increasing number of angry faces in an otherwise neutral crowd. Moreover, socially anxious individuals also avoid happy faces (Heuer et al., 2007; Roelofs, Putman, Schouten, Lange, van Peer, & Rinck, in press). In two experiments, we therefore examined whether CBM-I influences reflexive behavioural responses to emotional crowds, using the Approach–Avoidance Task (AAT) developed by Rinck and Becker (2007).

We predicted that negatively trained participants (NETs) would increasingly avoid stimuli of neutral–angry crowds with an increasing number of angry faces. For happy–angry crowds, we hypothesized that both emotional expressions would be seen as threatening by NETs and reacted with an overall avoidance response, irrespective of the ratio. Predictions for positively trained participants (POTs) were more difficult. As 'normally anxious' individuals are thought to hold a positive bias (e.g. Joormann & Gotlib, 2007), it could be argued that POTs' avoidance of increasingly angry crowds should be less pronounced or that they should start off with an approach tendency which weakens as the crowds become more negative. It is more likely, however, that POTs do not show any specific behavioural impulse as they may simply stick to task instructions (see Lange et al., 2008).

DOI: 10.1002/acp

EXPERIMENT 1

Method

Participants

Participants were preselected according to their trait–anxiety scores on the Dutch version of the State-Trait anxiety inventory (STAI; Van der Ploeg, Defares, & Spielberger, 1980) that 270 students had filled in at the beginning of their study year. In order to be able to induce and detect changes in anxiety, only students with scores around the mean (between 32 and 39) were invited to participate in the experiment. After exclusion of two participants due to technical problems, 68 second-year psychology students (88.2% female) of the University of Utrecht participated, 34 in the POT condition and 34 in the NET condition. The age of the participants ranged from 19 to 31 years ($M = 20.71$; $SD = 2.27$). An experimental session lasted about 1.5 hours and students received course credit.

Materials and measures

Interpretation training. The interpretation procedure made use of text materials that had been successfully employed earlier (e.g. Mathews & Mackintosh, 2000), in a Dutch translation utilized by Salemink, van den Hout and Kindt (2007a, b). In each of eight training blocks, participants read 13 text vignettes with descriptions of ambiguous social scenes. Each scene consisted of three sentences, which could be 'disambiguated' by filling in the missing letter of the last word. After this 'disambiguation,' the scenes were consistently either positive or negative, depending on the training condition. Eight of the 13 vignettes in each block were used for training purposes, but each block also contained three filler items and two probe items to monitor training effect across blocks. The filler items were added to obscure the direction of the training, and the probe items (one positive, one negative) were kept constant across the two training conditions. Differences in speed for resolving probe fragments served as manipulation check (Mathews & Mackintosh, 2000). Participants read the vignettes sentence by sentence (self-paced) on a computer monitor. When they reached the last word (-fragment), they filled in the missing letter. Then a comprehension question followed to enhance the interpretation given to the meaning of the scene.

Recognition task. In addition to the probe items that served as manipulation check throughout the training, another manipulation check consisted of 10 additional ambiguous social scenes, presented as before, with the following differences: A title was added, and the scene's resolution remained ambiguous, even upon completion of the word fragment. Next, participants were presented with each of the 10 scene titles and four possible interpretations, a positive one, a positive foil, a negative one and a negative foil. Participants rated every interpretation for its similarity to the original scene on a 4-point scale. Scores could range from (1) *very different in meaning* to (4) *very similar in meaning*.

Approach–avoidance task. The AAT was identical to the one used by Lange et al. (2008). A selection of 36 colour photos of 12 individuals (all male), each one presenting three different expressions, angry, neutral and happy, was taken from the Karolinska Directed Emotional Faces database (Lundqvist, Flykt, & Öhman, 1998). Matrices/crowds of 12 (4×3) facial expressions were constructed to vary in the degree of social approval/disapproval. Two types of crowds were created: Neutral–angry combinations and happy–angry combinations. The degree of threat was varied by gradually manipulating the ratio between pictures of the two target expressions of each crowd. Seven different ratios were

DOI: 10.1002/acp

composed: 12:0 (e.g. 12 neutral and zero angry pictures), 11:1, 9:3, 6:6, 3:9, 1:11 and 0:12. Each individual and emotional expression was randomly presented at any position. Every matrix was constructed in two different colour shadings (reddish, brownish) and in seven different sizes ranging from 200×202 to 760×768 pixels.

Procedure

When entering the laboratory, participants signed informed consent forms. Then they were seated approximately 50 cm from a computer monitor in a soundproof cubicle and asked to complete a first set of questionnaires: general screening questions, the Liebowitz social anxiety scale (LSAS; Liebowitz, 1987) and both versions of the STAI. Then they were randomly assigned to either positive or negative interpretation training and started the training program by pressing the keyboard's space bar. Each scene appeared on the screen one sentence at a time, and participants advanced to the next sentence until the word fragment was presented. As soon as they recognized the word, they pressed the space bar and then typed the first missing letter. Then the full word was shown, followed by the comprehension question. Feedback for key-press responses came in the form of a message with a blue or red background (for correct or incorrect responses, respectively). After training (45 minutes), participants filled in the STAI-State and the LSAS again and completed the recognition task (15 minutes; see Mathews & Mackintosh, 2000 for details).

Then, for the AAT, a standard computer joystick was located 25 cm between the participant and the monitor. Participants were asked to start each new trial by pressing the 'fire'-button. Then a medium sized crowd appeared on a black screen, and participants were instructed either to push or pull with the joystick, depending on the colour shading of the display (for details see: Lange et al., 2008). When the joystick was pulled, the display increased in size to give the impression of pulling the crowd closer. When pushing the joystick, the size of the display decreased in size to give the impression that the crowd was pushed away. Participants were instructed to move the joystick as quickly as possible, until the display disappeared. Reaction time (RT) measurement started upon presentation of the crowd and stopped at the end position when the crowd disappeared. The participant started the next trial by moving the joystick back to the central position and pushing the button. Participants were given 24 practice trials, before completing two blocks of 168 experimental trials each. The instructions concerning pulling and pushing the joystick, according to the shading of the display (brown or red), were counterbalanced across participants. At the end, participants were debriefed, compensated and thanked for their participation.

Results

Questionnaires

Before the start of the interpretive bias training, scores from participants in the positive and negative training group did not differ significantly on the STAI-Trait, $t(66) = 0.34$ ($M_{\text{overall}} = 36.03$, SD = 5.57), STAI-State, $t(66) = 0.77$ ($M_{\text{overall}} = 34.01$, SD = 6.78) and the LSAS social anxiety, $t(66) = 1.53$ ($M_{\text{overall}} = 30.81$, SD = 11.45). A 2 (training group) $\times$ 2 (time: before *vs*. after training) ANOVA to examine the effect of training revealed no significant effects for STAI-State and LSAS scores. In general, there was a main effect of time indicating that state anxiety and social anxiety decreased in both groups during the experiment irrespective of type of training, F (1, 66) = 4.39, $p = .02$ (state anxiety; $M_{\text{decrease}} = 1.82$) and F (1, 66) = 5.85, $p = .04$ (LSAS; $M_{\text{decrease}} = 1.19$).

DOI: 10.1002/acp

Manipulation checks

Reaction to probes. First, all probe trials with RTs ranging three standard deviations above or below the mean (5.9%) as well as error trials (10%) were removed from the data set. To measure whether the manipulation was effective, RTs in response to the probes were analysed with a 2 (direction of training: negative, positive) × 2 (probe valence: negative, positive) mixed-design ANOVA. The expected training × probe valence interaction was significant, $F(1, 66) = 14.11$, MSE = 17502.57, $p < .001$. As expected, POTs were faster when responding to *positive* probes than to negative probes ($M_{\text{Probe+}} = 1139$ ms, SD = 271 ms, *vs.* $M_{\text{Probe−}} = 1316$ ms, SD = 312 ms), but NETs reacted about equally fast to positive and negative probes ($M_{\text{Probe+}} = 1267$ ms, SD = 337 ms *vs.* $M_{\text{Probe−}} = 1273$ ms, SD = 306 ms).

Recognition task. A 2 (direction of training: negative, positive) × 2 (interpretation valence: negative, positive) × 2 (target: real, foil) mixed-design ANOVA was used to analyse the recognition task data. The predicted three-way interaction was significant, $F(1, 66) = 6.6$, MSE = 0.06, $p = .013$. *Post-hoc* analysis revealed that the training × valence interaction was significant both for real, $F(1, 66) = 26.0$, MSE = 0.18, $p < .001$, and foil sentences, $F(1, 66) = 17.1$, MSE = 0.09, $p < .001$. POTs generally tended to interpret social information more positively, NETs on the other hand did not interpret the information differently (for more details see Salemink, van den Hout, & Kindt, in press, because this subset of data was published in part in footnote 2.)

Approach–avoidance task

General AAT effects. First, all trials with RTs ranging three standard deviations above or below the mean (1.8%) as well as error trials (2.2%) were removed from the data set. After being identified as outliers regarding overall means as well as means for both crowd types separately, the data from six participants were excluded from further analysis. The data from another participant who did not follow instructions were excluded as well. Consequently, data of 61 participants were analysed, 29 in the POT and 32 in the NET condition.

AAT effects were calculated by subtracting each individual's mean RT for pulling a certain kind of crowd from the mean RT for pushing it. The resulting difference scores were entered into an ANOVA. There was a nonsignificant main effect of direction of training, $F(1, 59) = .69$, MSE = 5788.39, $p = .41$. A main effect for crowd type, $F(1, 59) = 22.35$, MSE = 6208.89, $p < .001$ ($M_{\text{Neutral–Angry}} = -9.88$, SD = 44.64 *vs.* $M_{\text{Happy–Angry}} = 15.65$, SD = 45.90) and a significant training × crowd type × expression ratio interaction, $F(6, 354) = 2.39$, MSE = 6010.86, $p = .028$ indicated that responses to the two crowd types were substantially different and that a separate analysis of the data within each crowd type would be necessary. Other effects in the overall design were nonsignificant, $ps > .24$.

AAT effects for neutral–angry crowds. When we first analysed the data from trials with neutral–angry crowds, all effects were nonsignificant, $ps > .13$.

Because possibly small effects might become diluted in the seven levels of expression ratio, the overall response pattern as reflected in the corrected gradients of each group's regression line was considered a more sensitive measure. When we examined these gradients, the number of angry faces in an otherwise neutral crowd appeared to play a different role for NETs than for POTs: The interaction between expression ratio and training was significant $F(1, 59) = 5.92$, MSE = 5368.11, $p = .02$. Regression lines were

DOI: 10.1002/acp

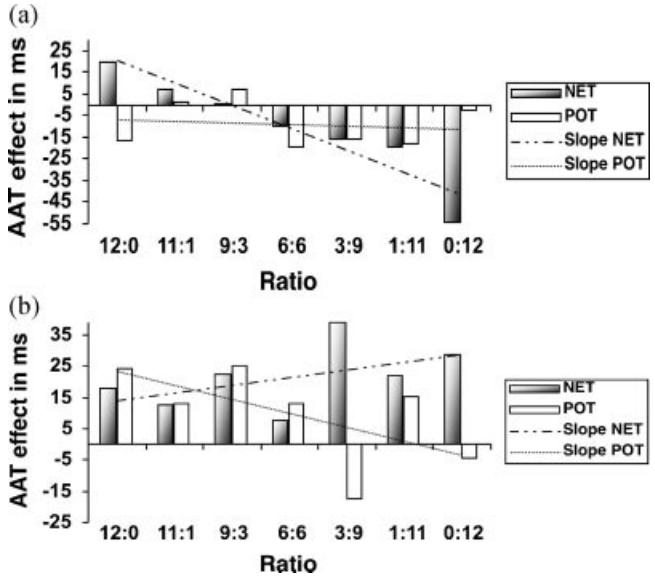

Figure 1. Mean AAT effects in Experiment 1 for: (a) neutral–angry crowds per ratio and training group, and (b) happy–angry crowds, where POT refers to positive and NET to negative training. The ratio of angry faces increases from left to right. For example, 11:1 refers to 11 neutral or happy faces and one angry face. AAT effect scores are calculated by subtracting RTs for pulling from RTs for pushing the joystick; negative scores refer to more 'avoidance' and positive scores refer to more 'approach'.

then analysed separately for each direction of training. The slope of NETs differed significantly from zero, $F(1, 28) = 11.88$, MSE = 6733.80, $p = .002$, but the slope for the POTs was not, $F(1, 31) = 0.36$, MSE = 4134.59, $p = .55$. Only the NETs became more avoidant as the number of angry faces in a neutral–angry crowd increased (Figure 1a).

AAT effects for happy–angry crowds. For the happy–angry crowds, the main effects of direction of training expression ratio and their interaction were nonsignificant, $ps > .16$. In contrast to the pattern for neutral–angry crowds, analyses of the slopes revealed no significant differences between NETs and POTs, $ps > .06$. Training did not change participants' reflexive behaviour towards happy–angry crowds (Figure 1b).

Discussion

The results of Experiment 1 suggest that (1) we were successful in inducing a positive and negative interpretation bias and that (2) an induced negative interpretation bias can potentially influence subsequent reflexive behaviours, depending on the types of emotional expressions combined in multi-facial displays. With an increasing number of angry faces in a neutral crowd, NETs became faster in pushing the crowds away (avoidance) than in pulling them closer (approach) similar to socially anxious individuals (see Lange et al., 2008). In contrast, POTs' pushing was about as fast as their pulling in all stimulus configurations. In the happy and angry face combination, no group differences in reflexive behaviour occurred.

A major concern, however, is the absence of any avoidance response of NETs to 0:12 happy–angry crowds. These are in fact 'all-angry' crowds and conceptually the same as the 0:12 neutral–angry crowds. The picture configuration with the 12 actors within a crowd

DOI: 10.1002/acp

was randomized for every new trial and the allocation of these all-angry crowds to one of the two crowd types was arbitrary. This means, though, that all-angry crowds, even if identical only by chance, were seen twice as frequently as the other configurations. As the order of trials was pre-randomized and therefore identical for each participant, it is possible that the specific positions of the neutral–angry 'all-angry crowd' in the program script may have had different effects on participants' responses than the positions of their happy–angry equivalents. To address these problems, we repeated the Experiment 1 with some minor changes.

EXPERIMENT 2

Method

Participants

The same selection criteria as in Experiment 1 led to the inclusion of 39 students (81.82% female) from Radboud University Nijmegen; 20 were randomly assigned to POT and 19 to NET. The age of the students ranged from 17 to 35 years ($M = 20.98$; SD = 3.01).

Materials and procedure

Materials and procedure were the same as in Experiment 1, with the following exceptions. The AAT program was revised such that the former 0:12 happy–angry crowds appeared where 0:12 neutral–angry stimuli were presented, and *vice versa*. Accordingly, we could investigate whether the differences in participants' responses to these stimulus categories may have been caused by an order effect or by subtle visual differences between these two categories.

Results

Questionnaires

Before the start of the interpretive bias training, participants in the positive and negative training group did not differ significantly for scores on the STAI-Trait, $t(32) = 0.42$ ($M_{\text{overall}} = 34.67$, SD = 4.47), STAI-State, $t(32) = 0.9$ ($M_{\text{overall}} = 31.13$, SD = 4.49) and the LSAS social anxiety, $t(32) = 1.36$ ($M_{\text{overall}} = 35.50$, SD = 15.08). A 2 (training group) × 2 (time: before *vs*. after training) ANOVA to examine the effect of training revealed no significant effects for STAI-State and LSAS scores, $ps > .40$.

Manipulation checks

Reaction times to probes. Before analysis, all trials (7.4%) with RTs three standard deviations above or below the mean were removed as well as all error trials were also removed (11.4%). Five subjects with more than 25% of errors were excluded from all analyses because the training may not have worked properly, leaving 18 participants in the POT condition and 16 participants in the NET condition.

When analysing the remaining data, the predicted training × probe valence interaction was significant, $F(1, 32) = 9.39$, MSE = 97930.93, $p = .004$: POTs were slower in responding to negative word fragments ($M = 1765$, SD = 491) than were NETs ($M = 1316$, SD = 397). The training conditions did not differ significantly in responding to positive probe trials, $M = 1407$, SD = 350 *vs*. $M = 1424$, SD = 290, respectively.

DOI: 10.1002/acp

The recognition task. The training × target (possible *vs*. foil) interaction effect was significant, $F(1, 32) = 21.09$, MSE = 0.281, $p < .001$, reflecting that POTs gave higher similarity ratings to possible interpretations than did NETs, but no differences occurred in response to foil interpretations. Also, the crucial group × valence interaction effect was significant, $F(1, 32) = 4.59$, MSE = 0.104, $p = .04$. POTs ($M = 2.46$, SD = 0.38) made more positive interpretations than NETs ($M = 2.17$, SD = 0.31), while NETs endorsed the negative interpretations ($M = 2.37$, SD = 0.39) more than the POTs ($M = 1.83$, SD = 0.24). Thus, participants had the tendency to interpret new ambiguous information with the valence they had been trained with.

Approach–avoidance task

Again, error trials (1.7%) as well as RTs ranging beyond three standard deviations above or below the mean (1.6%) were removed. In the overall analysis of the AAT-effect scores, no effect was significant, $ps > .17$. Specifically, the relevant crowd type × angry faces × training was not significant, $F(6, 192) = 0.92$, MSE = 3731.56, $p = .48$.

Visual inspection of Figure 2 reveals that, independent of crowd type, the AAT-scores of POTs decreased as more angry faces were in a crowd. But an explorative slope analysis only showed a marginal significant interaction of training × number of angry faces in the neutral–angry crowds, $F(1,32) = 3.23$, MSE = 3791.14, $p = .08$. When regression lines were analysed separately for each direction of training, neither slope differed significantly from zero; NET: $F\ (1, 15) = 0.943$, MSE = 4130.96, $p = .35$; POT: $F\ (1, 17) = 2.66$, MSE = 8682.31, $p = .12$. The interaction term for the slopes in happy–angry crowds did not approach significance, $F(6, 192) = 1.4$, MSE = 3641.86, $p = .23$. Therefore, no further exploration was undertaken.

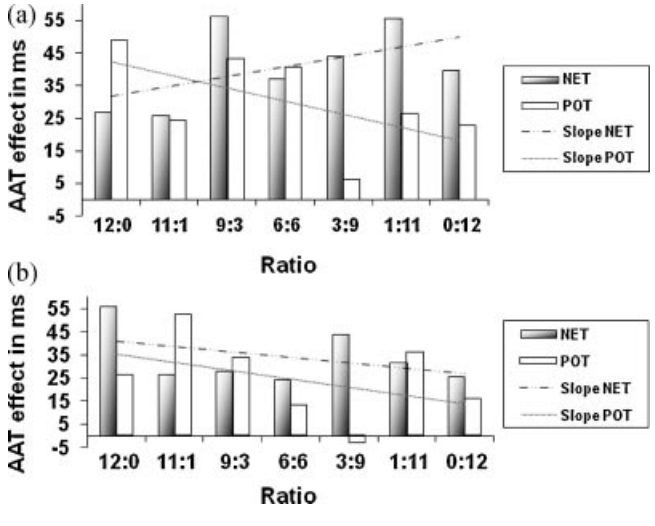

Figure 2. Mean AAT effects in Experiment 2 for: (a) neutral–angry crowds per ratio and training group, and (b) happy–angry crowds, where POT refers to positive and NET to negative training. The ratio of angry faces increases from left to right. For example, 11:1 refers to 11 neutral or happy faces and one angry face. Negative scores refer to more 'avoidance' and positive scores refer to more 'approach'.

DOI: 10.1002/acp

The AAT-scores concerning all-angry crowds in the happy–angry condition appeared comparable between the two experiments irrespective of the fact that the order of the stimuli was altered in Experiment 2. Responses to the all-angry crowds in the neutral–angry condition now seemed to show a similar pattern as in the happy–angry conditions.

Discussion

Experiment 2 could not substantiate the claim that an induced (negative) interpretation bias influences subsequent reflexive behaviour impulses. Although the CBM-I procedure successfully changed biases as it did in Experiment 1, there was no transfer to the AAT. Visual inspection of the data suggests that the patterns of responses to both crowd types differ quite substantially across experiments. In Experiment 2, it seems as if the POTs were becoming more avoidant with every additional angry face. It also appears as if POTs show more pronounced approach to all-neutral crowds; with a positive bias these particular crowds may appear more approachable. However, in light of the results for some of the other expression ratios, and considering the absence of a significant interaction, this interpretation is highly speculative. Unlike Experiment 1, the response patterns to 0:12 neutral–angry crowds and 0:12 happy–angry crowds were similar in Experiment 2.

GENERAL DISCUSSION

In two experiments, the transfer of an interpretative bias modification procedure to reflexive approach–avoidance tendencies towards facial crowds was investigated. First, a selection of medium range trait anxious individuals underwent training to endorse either positive or negative interpretations of social scenarios. Second, as an analogue for approach and avoidance tendencies, participants were instructed to pull or push stimulus pictures denoting different ratios of neutral–angry or happy–angry faces by means of a joystick. The experiments were designed to test whether the mere presence of a negative interpretation bias as observed in socially anxious individuals is related to reflexive avoidance patterns towards emotional faces also seen in socially anxious individuals.

Results of the AAT in Experiment 1 suggested an increasing degree of avoidance for the NETs as the number of angry faces increased in neutral–angry crowds (Lange et al., 2008). However, NETs' responses to all-angry stimuli randomly assigned to the neutral–angry category differed considerably from responses to the equivalent pictures assigned to the happy–neutral category, even though they were conceptually the same. Experiment 2 was performed to replicate these findings and to clarify the role of possible order-effects of all-angry crowds in the experimental sequence of the AAT. The results of the AAT, however, were not in line with the predictions: CBM-I had no impact on reflexive behaviour.

Baseline and effect of training

There is some debate about whether evidence for CBM-I requires a full crossover in the form of POTs reacting faster to positive than to negative probes and NETs reacting faster to negative than to positive probes. Seemingly, the results of our manipulation checks imply that only participants who received *positive* interpretation training responded accordingly. NETs responded equally fast to positive and negative probes. It would have been more elegant to measure interpretation biases (and AAT scores) before the training to establish a

DOI: 10.1002/acp

baseline and determine the degree of change due to the training. Salemink et al. (in press), on the other hand, did not use a pre–post comparison but assigned an independent control group of average anxious participants to complete the recognition task. Their between-group analyses revealed that, compared to the control group, interpretations in *both* training groups were successfully modified according to the direction of the training (Salemink et al., in press). Because the pattern of our training results are comparable to that of Salemink et al., the response pattern of their untrained participants may serve as an indication that the training in the present study might have worked both ways too. In addition, our study was meant as an analogy to cognitive biases in social anxiety, therefore, it is also noteworthy that when investigating the manifestation of interpretation biases in socially (non-) anxious populations, Hirsch and Mathews (2000) confirmed that nonanxious controls indeed reacted faster to positive than to negative probes, whereas socially anxious individuals reacted equally fast to both probes.

The research literature also indicates that nonanxious or averagely anxious participants may have a positive interpretation bias 'protecting' them from psychopathology (Garner, Mogg, & Bradley, 2006). Therefore, it is plausible that the NETs in the present study were deprived of their formerly positive bias. This possibility could explain why they did not show significant RT differences for positive compared to negative probes or why they did not show significant recognition differences for negative as compared to positive interpretation options. POTs did show these differences and it appeared as if *they* were the ones profiting from the training, even though they might simply have been doing what they always do: prefer positive interpretations.

CBM-I and generalization

From the present results it can be concluded that transfer from CBM-I procedures to different bias measures, cognitive domains and behaviour remains challenging. The results of Experiment 1 were promising until they were not replicated in Experiment 2. These difficulties are in line with inconsistencies of earlier findings. Though, for example, Mackintosh et al. (2006) showed that interpretation training transferred from auditory to visual material, and that training in one context can lead to biased interpretations in another, Salemink et al. (2007a), on the other hand, did not find any transfer of training to an extrinsic affective Simon task and to open-ended questions on ambiguous vignette continuations. In another experiment, Salemink et al. (in press) showed that the training transferred to vignette recognition in another domain (academic performance) but not to other tasks such as a paper–pencil interpretation bias measure or emotionality rating if one were confronted with evaluative comments of an actor in video clips (Amir, Beard, & Bower, 2005). By using a different kind of interpretive bias training, however, Wilson, MacLeod, Mathews, and Rutherford (2006) did show transfer to a similar video task.

Benign interpretation training has the potential to reduce negative biases in (clinically) anxious individuals (e.g. Murphy, Hirsch, Mathews, Smith, & Clark, 2007), but it is unclear whether the training has potential disorder-specific effects on behaviour patterns. Concerning the transfer of CBM-I to behaviour, to our knowledge only one study found tenuous evidence that a positive CBM-I in spider phobic individuals may lead to changes in approach behaviour (Teachman & Addison, 2008). Salemink et al. (in press) argued that the lack of ambiguity in the stimuli with which transfer of interpretation bias is to be measured seems crucial. It is possible that intermediate presentations of unambiguous stimuli may counteract former bias manipulations. Taking this argument into account, it

DOI: 10.1002/acp

remains unclear whether facial expressions are ambiguous enough to allow a biased interpretation to come into effect. In addition, the neurological hard-wiring of facial emotion detection (Vuilleumier, 2002) makes cognitive influences even more unlikely. Surprisingly, then, one would assume that negative (but also positive) interpretations would act the most on behavioural responses to neutral faces, but as has been shown by Lange et al. (2008), this is not the case.

Finally, it is plausible to assume that avoidance behaviours to happy and angry, but not to neutral faces (e.g. Roelofs et al., in press), are either not driven by a cognitive bias as the one measured in socially anxious individuals and induced by CBM-I procedures, or provide an inadequate measure for this kind of cognitive bias. This could mean that a negative interpretation bias as measured in socially anxious individuals and their avoidance tendencies in response to happy and angry faces are independent phenomena not influencing each other. If this is true, then in order to find behavioural changes as consequence of CBM-I, behaviours directly related to cognitive biases must be chosen.

In sum, we successfully induced an interpretation bias, although our results could not unequivocally provide evidence that CBM-I potentially transfer to reflexive anxiety-related behaviour. In fact, more questions are raised than answered, as Experiment 1 seems to show some evidence of such a transfer while Experiment 2 did not. In the future, it seems crucial to discover causal links between biased interpretation and (indirectly) observable behaviour. These links are necessary for claiming that, rather than merely being a symptom, cognitive biases are initiating factors in bringing about social anxiety.

ACKNOWLEDGEMENTS

This research was funded by grants from NWO (Netherlands Organization for Scientific Research) and BSI (Behavioural Science Institute). The authors would like to thank Yora Overes for her help for collecting the data and Rinske de Graaff-Stoffers and Oliver Langner for their statistical advice and support.

REFERENCES

Amir, N., Beard, C., & Bower, E. (2005). Interpretation bias and social anxiety. *Cognitive Therapy and Research*, *29*, 433–443.

Beard, C., & Amir, N. (2008). A multi-session interpretation modification program: Changes in interpretation and social anxiety symptoms. *Behaviour Research and Therapy*, *46*, 1135–1141.

Clark, D. M., & Wells, A. (1995). A cognitive model of social phobia. In R. Heimberg, M. Liebowitz, D. Hope, & F. Schneier (Eds.), *Social phobia diagnosis, assessment, and treatment* (pp. 69–112). New York: Guilford Press.

Foa, E. B., Franklin, M. E., & Kozak, M. J. (2001). Social phobia: An information-processing perspective. In S. G. Hofmann, & P. M. DiBartolo (Eds.), *From social anxiety to social phobia: Multiple perspectives* (pp. 268–280). Needham Heights, MA: Allyn & Bacon.

Garner, M., Mogg, K., & Bradley, B. P. (2006). Fear-relevant selective associations and social anxiety: Absence of a positive bias. *Behaviour Research and Therapy*, *44*, 201–217.

Heuer, K., Rinck, M., & Becker, E. S. (2007). Avoidance of emotional facial expressions in social anxiety: The approach–avoidance task. *Behaviour Research and Therapy*, *45*, 2990–3001.

Hirsch, C. R., & Mathews, A. (2000). Impaired positive inferential bias in social phobia. *Journal of Abnormal Psychology*, *109*, 705–712.

DOI: 10.1002/acp

Huppert, J. D., Pasupuleti, R. V., Foa, E. B., & Mathews, A. (2007). Interpretation biases in social anxiety: Response generation, response selection, and self-appraisals. *Behaviour Research and Therapy*, *45*, 1505–1515.

Joormann, J., & Gotlib, I. H. (2007). Selective attention to emotional faces following recovery from depression. *Journal of Abnormal Psychology*, *116*, 80–85.

Lange, W.-G., Keijsers, G. P. J., Becker, E. S., & Rinck, M. (2008). Social anxiety and evaluation of social crowds: Explicit and implicit measures. *Behaviour Research and Therapy*, *46*, 932–943.

Liebowitz, M. R. (1987). Social phobia. *Modern Problems of Pharmacopsychiatry*, *22*, 141–173.

Lundqvist, D., Flykt, A., & Öhman, A. (1998). *The Karolinska Directed Emotional Faces—KDEF*, CD ROM from Department of Clinical Neuroscience, Psychology section, ISBN 91-630-7164-9. Stockholm Sweden: Karolinska Institutet.

Mackintosh, B., Mathews, A., Yiend, J., Ridgeway, V., & Cook, E. (2006). Induced biases in emotional interpretation influence stress vulnerability and endure despite changes in context. *Behavior Therapy*, *37*, 209–222.

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

Merckelbach, H., Van Hout, W., Van den Hout, M. A., & Mersch, P. P. (1989). Psychophysiological and subjective reactions of social phobics and normals to facial stimuli. *Behaviour Research and Therapy*, *27*, 289–294.

Murphy, R., Hirsch, C. R., Mathews, A., Smith, K., & Clark, D. M. (2007). Facilitating a benign interpretation bias in a high socially anxious population. *Behaviour Research and Therapy*, *45*, 1517–1529.

Philippot, P., & Douilliez, C. (2005). Social phobics do not misinterpret facial expression of emotion. *Behaviour Research and Therapy*, *43*, 639–652.

Rinck, M., & Becker, E. S. (2007). Approach and avoidance in fear of spiders. *Journal of Behavior Therapy and Experimental Psychiatry*, *38*, 105–120.

Roelofs, K., Putman, P., Schouten, S., Lange, W.-G., van Peer, G., & Rinck, M. (in press). Gaze direction affects approach–avoidance behavior to angry faces and not to happy faces. *Behaviour Research and Therapy.*

Salemink, E., van den Hout, M., & Kindt, M. (2007a). Trained interpretive bias: Validity and effects on anxiety. *Journal of Behavior Therapy and Experimental Psychiatry*, *38*, 212–224.

Salemink, E., van den Hout, M. A., & Kindt, M. (2007b). Trained interpretive bias and anxiety. *Behaviour Research and Therapy*, *45*, 329–340.

Salemink, E., van den Hout, M., & Kindt, M. (in press). Generalization of modified interpretive bias across tasks and domains. *Cognition & Emotion*, 1–17.

Teachman, B. A., & Addison, L. M. (2008). Training non-threatening interpretations in spider fear. *Cognitive Therapy and Research*, *32*, 448–459.

Van der Ploeg, F. A., Defares, P. B., & Spielberger, C. D. (1980). *Handleiding bij de Zelf Beoordelings Vragenlijst, ZBV.* Lisse, The Netherlands: Swets & Zeitlinger.

Voncken, M. J., Bögels, S. M., & de Vries, K. (2003). Interpretation and judgmental biases in social phobia. *Behaviour Research and Therapy*, *41*, 1481–1488.

Vuilleumier, P. (2002). Facial expression and selective attention. *Current Opinion in Psychiatry*, *15*, 291–300.

Wilson, E. J., MacLeod, C., Mathews, A., & Rutherford, E. M. (2006). The causal role of interpretive bias in anxiety reactivity. *Journal of Abnormal Psychology*, *115*, 103–111.

DOI: 10.1002/acp