1 **Angry-Happy Interpretations of Ambiguous Faces in Social Anxiety Disorder**

2 Running Header: Angry-Happy Interpretation in SAD

3 Keren Maoz ^a

4 Sharon Eldar ^a

5 Joel Stoddard ^b

6 Daniel S. Pine ^c

7 Ellen Leibenluft ^b

8 Yair Bar-Haim ^{a,d}

9

10 ^a School of Psychological Sciences, Tel Aviv University, Tel Aviv, Israel

11 ^b Section on Bipolar Spectrum Disorders, Emotion and Development Branch, National
12 Institute of Mental Health, National Institutes of Health, Department of Health
13 and Human Services, Bethesda, Maryland, USA.

14 ^c Section on Development and Affective Neuroscience, Emotion and Development
15 Branch, National Institute of Mental Health, National Institutes of Health,
16 Department of Health and Human Services, Bethesda, Maryland, USA.

17 ^d The Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, Israel

18

19 **Corresponding Author:** Correspondence concerning this article should be addressed
20 to Keren Maoz, School of Psychological Sciences, Tel Aviv University, Ramat Aviv,
21 Tel Aviv 69987, Israel. E-mail: kerenmao@mail.tau.ac.il Phone number: 972-54-
22 2293397

23

24

Abstract

Social Anxiety Disorder (SAD) is characterized by a tendency to interpret ambiguous social cues as negative. Here we tested whether interpretation of ambiguous faces differs between participants with SAD and non-anxious controls. Twenty-seven individuals with SAD and 21 non-anxious control participants completed an emotion recognition task in which they judged ambiguous morphed faces as happy or angry. Participants with SAD judged a higher proportion of the faces as angry compared to non-anxious participants, and were slower to judge faces as angry compared to happy, while no such reaction time bias manifested in the control group. Finally, happy judgments were slower in the SAD group compared to the control group, while angry judgments were faster in the SAD group compared to the control group. These findings provide evidence for a negative bias in resolving emotional ambiguity in facial expressions among individuals with SAD.

Keywords: Social phobia, Interpretation Bias, Ambiguous facial expressions, Emotion, Anger

1. Introduction

Cognitive models suggest that social anxiety disorder (SAD) is associated with cognitive biases that promote and maintain this condition, including a tendency to interpret ambiguous social cues as negative (Clark, 2001; Clark and Wells, 1995; Rapee and Heimberg, 1997). Experimental studies have also demonstrated a causal association between negative interpretation bias and anxiety reactivity (Mathews and Mackintosh, 2000; Wilson et al., 2006). Research in interpretation biases in social anxiety can be broadly divided into two lines of studies. The first includes studies on interpretation of socially relevant verbal stimuli, such as descriptions of ambiguous social scenarios (e.g., Amir et al., 1998b; Huppert et al., 2003; Muris et al., 2000; Stopa and Clark, 2000; Voncken et al., 2003), or ambiguous sentences with social content (Beard and Amir, 2009; Huppert et al., 2007). In socially anxious vs. non-anxious participants, these studies generally find a bias towards negative interpretation of ambiguous, socially relevant information. The second line of studies focuses on biased interpretation of facial expressions of emotion. Processing facial expressions of emotion is central to understanding and evaluating social situations, and is thought to be biased in individuals with SAD who are typically preoccupied with others' evaluation or scrutiny. Negative interpretations of ambiguous facial expressions may result in a tendency to see others as more criticizing, threatening, or hostile, which may contribute to elevated anxiety in social situations.

Recently, studies have begun to use computerized morphing procedures to generate systematic ambiguity in facial expressions and examine biases in emotion identification and classification among socially anxious individuals. Some studies have used sequences ranging from a neutral expression to an emotional expression, examining individual differences in the level of intensity needed to identify an

emotion. This work has generated mixed results, with some studies finding a negativity bias in socially anxious vs. non-anxious participants (Gilboa-Schechtman et al., 2008; Heuer et al., 2010; Joormann and Gotlib, 2006; Yoon et al., 2014), while others fail to find this pattern (Philippot and Douilliez, 2005). Since these studies use face stimuli that vary in their emotional intensity (e.g., 50% anger vs. 20% anger), the results may reflect a general sensitivity to emotional cues, rather than interpretative biases (Jusyte and Schönenberg, 2014). It has been suggested that ambiguous faces that contain conflicting information (e.g., morphing an angry and a happy expression), may be more effective in eliciting an interpretation bias because they create a conflict in the classification of ambiguous expressions. This method has been utilized by a smaller number of studies (e.g., Garner et al., 2009; Jusyte and Schönenberg, 2014; Reeb-Sutherland et al., 2015; Richards et al., 2002) and again, the results are mixed. In socially anxious vs. non-anxious participants, two studies found enhanced interpretation of fear (Reeb-Sutherland et al., 2015; Richards et al., 2002); one study found lower sensitivity for fear (Garner et al., 2009); and one study found no between-group difference (Jusyte and Schönenberg, 2014).

Surprisingly, none of these studies found interpretation biases related to anger among socially anxious participants compared to non-anxious participants. Of note anger is the emotion most directly associated with social threat and in SAD samples, angry faces induce attention biases (Gilboa-Schechtman et al., 1999; Mogg et al., 2004), identification sensitivity differences (Joormann and Gotlib, 2006), and enhanced amygdala activation (Stein et al., 2002). Accordingly, it is expected that anger expressions would play an important role in cognitive processes related to SAD. Two previous studies that compared non-clinical samples of individuals with high and low social anxiety found less positive interpretation in the high-anxiety group of

schematic or composite expressions that combined happy-angry conflicting information (i.e., a smiling mouth with angry eyes) (Coles et al., 2008; Gutiérrez-García and Calvo, 2014). The current study tests for the presence of negative interpretation bias in clinical SAD patients compared to non-anxious participants. We specifically focused on the interpretation of anger in ambiguous faces containing conflicting information in the form of happy-angry morphed faces.

Two previous studies have examined interpretation of angry-happy morphed faces in clinical SAD samples (Garner et al., 2009; Jusyte and Schönenberg, 2014) and found no difference in interpretation patterns of angry-happy morphed faces between SAD and non-anxious control participants. The failure of these two studies to find SAD-related differences in the interpretation of angry-happy morphed faces may be related to their applied methodology. Specifically, in both studies participants were requested to rate the intensity of the emotion perceived following each morphed face presentation. Evaluating facial emotion on an intensity scale may have required participants to engage in a more thoughtful process of evaluation compared with the more basic and simple process of distinguishing between angry and happy faces. Thus, in the current study we used a simple two forced choice task with brief presentations in order to encourage basic interpretation processes rather than higher level intensity evaluation processes. Moreover, Jusyte and Schönenberg (2014) used happy endpoints that consisted of open mouth smiles but angry endpoints that consisted of closed mouth expressions. According to these authors, this difference may have resulted in biased judgment due to the different pattern of teeth exposure. They suggest that one way to avoid artifacts in happy/angry morphs would be to choose models that display both happy and angry expressions with visible teeth. In

the current study we followed this rationale selecting both the happy and angry endpoints of the morphed sequences to have open mouths.

Notably, most of the existing clinical studies quantified interpretation biases based on participants' response types, such as the number of negative responses (Jusyte and Schönenberg, 2014), emotional intensity ratings (Jusyte and Schönenberg, 2014; Philippot and Douilliez, 2005), emotional intensity for correct identification (Joormann and Gotlib, 2006), accuracy rates (Philippot and Douilliez, 2005), or signal-detection based indices (Garner et al., 2009). Another aspect of biased interpretation processing may be reflected in the speed of interpretation. Specifically, if negative interpretations are more accessible to the anxious individual's mind (Amir et al., 1998a; Amir et al., 1998b; Stopa and Clark, 2000), then negative interpretations will be carried out faster than positive or neutral interpretations. This pattern has been demonstrated with non-clinical high vs. low socially anxious individuals (Gutiérrez-García and Calvo, 2014). Here, in addition to measuring the percent of negative interpretations in participants with SAD and non-anxious controls, we also tested the speed of negative vs. positive interpretations.

The identification of biased interpretation of negative stimuli in SAD has inspired clinical translational attempts targeting bias modification. So far these translational attempts have relied on verbal descriptions of social scenarios or sentences, with several studies demonstrating that promoting benign or positive interpretations (e.g., with feedback or repeated exposure to positive or benign resolutions of ambiguity) is associated with reduced anxiety levels (e.g., Amir and Taylor, 2012; Beard and Amir, 2008; Murphy et al., 2007). To our knowledge no SAD intervention study targeted biased interpretation of facial expressions. The first step toward such translation is to demonstrate a measurable interpretation bias in SAD

using a task that can later be used for modification of this specific bias. The aim of the current study was to document an interpretation bias in socially anxious individuals (compared to control participants) using a paradigm appropriate for future CBM intervention. We thus used a variant of a task that has been applied to modify biased interpretation of angry faces in the context of aggressive behavior and mood disturbances (Penton-Voak et al., 2013; Stoddard et al., 2016). In these studies, participants first completed a measurement task in which they judged whether morphed faces were happy or angry. The second phase included cognitive training in which participants received systematic feedback after each response, designed to shift their judgements toward more positive interpretation of the ambiguous morphed faces. Here we used a similar measurement task to examine whether individuals with SAD show biased interpretation of morphed faces compared to controls. We reasoned that, if SAD-related individual differences emerged on this measurement task, future translation toward an interpretation bias modification protocol would be facilitated, given the success of previous studies in modifying this specific bias using systematic feedback.

In sum, we tested whether participants with SAD demonstrate interpretation biases when identifying the emotional content of ambiguous facial stimuli. We used a task similar to that in cognitive bias modification (CBM) studies (Penton-Voak et al., 2013; Stoddard et al., 2016). Specifically, we used a two forced-choice emotion recognition task, with brief presentations of morphed pictures created by blending angry and happy expressions of the same individuals. Two hypotheses were examined. First, we expected participants with SAD to interpret a higher proportion of the faces as angry, relative to non-anxious control participants. Second, we expected

participants with SAD to make angry interpretations faster than happy interpretations. This reaction time (RT) bias was not expected in the non-anxious control group.

2. Method

2.1. Participants

The SAD group consisted of 27 participants (mean age=28.4 years, SD=6.8; 14 females). These participants were recruited from the community as part of an ongoing anxiety treatment program run at the University. Free treatment as part of efficacy research is offered and participants agreed to complete our task at baseline before treatment. This procedure was approved by the Institutional Review Board and data are protected through a standardized de-identification protocol. Severity of social anxiety symptoms was assessed using the Liebowitz Social Anxiety Scale (LSAS; Liebowitz, 1987). This scale consists of 24 items describing social interactions and performance situations. The LSAS possesses strong psychometric properties (Fresco et al., 2001). The Hebrew version of the LSAS was found valid and reliable (Levin et al., 2002). Clinical diagnosis of SAD was established by a trained clinical psychologist using a structured interview (Mini International Neuropsychiatric Interview, MINI, Sheehan et al., 1998) , and relying on a LSAS score that is greater than 60, which has been established as a reliable cutoff for generalized SAD (Mennin et al., 2002; Rytwinski et al., 2009).

The control participants were recruited through advertisement in social media, and consisted of 21 participants (mean age=25.5 years, SD=3.5; 12 females) who scored 30 or less on the LSAS. A cutoff score of 30 was found to provide the best

balance between false positive and false negative diagnostic errors in classifying individuals with and without social anxiety disorder (Mennin et al., 2002; Rytwinski et al., 2009).

The study was approved by the University's Institutional Review Board and all participants provided signed informed consent.

2.2. The emotion recognition task

Stimuli for the task were generated using Morpheus Photo Morpher v3.16. Four sequences of morphed faces were generated based on the happy and angry pictures of two males and two females taken from the NimStim set (Tottenham et al., 2009; models 10, 18, 37, 41). The faces were selected to represent both genders and different racial origins (Caucasian, African-American, Hispanic, and Asian), and following pilot tests that demonstrated test-retest validity for these morphed sequences in the emotion perception task. Each sequence consisted of 15 faces equally-spaced on a continuum between the happy and angry endpoints (Figure 1). Each face from each morphed sequence was presented three times, for a total of 180 trials (4 sequences X 15 faces X 3 repetitions). Faces were displayed in a random order.

Each trial began with a fixation cross (800-1200 ms), followed by a color morphed face picture (90mm in height and 70mm in width). The face was displayed for 200ms, and then masked by a scrambled face display for 200ms. Then, a question mark appeared on the screen and remained until a response was made. Participants were instructed to press one of two designated buttons as fast as they could to indicate whether the face was "angry" or "happy". Following the participant's response the next trial began 100-400 ms later. The exact ISI was determined randomly.

Two measures were derived from the task: a) the percent of "angry" responses (the total number of angry responses divided by the total number of trials) was used as an index of participants' tendency to interpret the face stimuli as negative; and b) mean reaction time (RT) for angry and happy interpretations were used to examine differential response time patterns. For illustration purposes we also calculated an RT bias score for each participant by subtracting their mean RTs to angry-interpreted faces from their mean RTs to happy-interpreted faces. Thus, a positive bias score reflect a tendency to make negative interpretations faster than positive interpretations, whereas a negative bias score reflects the opposite tendency.

2.3. Procedure

As mentioned above, participants from the SAD group were recruited as part of an ongoing anxiety treatment program and completed the task after their diagnosis was confirmed and before treatment began. The non-anxious control group was recruited from advertisements. Low LSAS scores were ascertained for this group. The lab session was identical for all participants. Participants arrived at the lab and signed an informed consent. They were seated approximately 80 cm from the screen and were told they were about to complete an emotion recognition task. They were instructed that, in each trial, a fixation cross will appear, followed by a brief presentation of a face. They were asked to indicate whether the face was angry or happy by pressing one of two labeled keyboard buttons. They were instructed to respond quickly while avoiding errors. This procedure took approximately 10 minutes.

3. Results

3.1. Data cleaning

To gauge overall engagement and compliance with task demands, we required that subjects achieve 70% accuracy when identifying the two overtly angry and two overtly happy facial expressions i.e., the two most extreme faces from each end of the morphed sequence of each set (Stoddard et al., 2016). Since each face was presented 3 times during the task, accuracy was based on a total of 48 trials (4 extreme faces x 4 sets x 3 repetitions). All participants correctly identified more than 70% of the faces, thus none of the data were removed from analysis.

For the percent of angry responses index no data cleaning was applied. For RT-based analyses, trials with RTs longer than 2000ms were excluded. Then, for each participant, trials with RTs deviating by more than 3 standard deviations from the mean of each response type (happy/angry) were also excluded. This resulted in the removal of 3% of all trials.

Means and SDs for participants' characteristics by group are provided in Table 1. There were no significant differences in age or gender between the SAD and the control group, $p>0.05$. As expected given the selection procedure, participants with SAD scored significantly higher on the LSAS than the non-anxious participants $t_{46}=20.89, p<0.001$, Cohen's $d=6.08$.

3.2. Judgments analysis

Participants with SAD judged a higher proportion of the faces as angry compared to the non-anxious participants, $t_{46}=2.78, p<0.05$, Cohen's $d=0.81$ (Figure 2A). We also tested group differences for the point at which judgments switch from predominantly happy to predominantly angry (i.e., the balance point, or 'category

boundary'). Following Pollak and Kistler (2002) we calculated the category boundary from 4-parameter logistic curves fitted to individual-level data. This analysis yielded a between-group effect, such that participants with SAD had a lower balance point compared to the control participants ($M=6.73$, $SD=0.88$ and $M=7.33$, $SD=0.63$, respectively, $t_{46}=2.66$, $p<0.05$, Cohen's $d=0.77$). This further corroborated our primary analysis by indicating that participants with SAD required less angry affect in morphs to switch their judgments from predominantly happy to predominantly angry.

3.3. Response time analysis

Raw RT means and standard deviations for angry and happy judgements by group are presented in Table 1. Mean RTs were submitted to a 2 x 2 ANOVA with Group (SAD, Control) as a between subjects factor and Judgment (happy, angry) as a repeated within-subject factor. The main effect of Judgment was not significant, $F(1,46)=3.13$, $p>0.08$, $\eta^2_p=0.06$. A main effect of group indicated that non-anxious control participants were overall faster than participants with SAD, $F(1,46)=4.9$, $p<0.05$, $\eta^2_p=0.10$. This main effect was qualified by a significant Group-by-Judgment interaction effect, $F(1,46)=6.87$, $p<0.05$, $\eta^2_p=0.13$. Follow-up contrasts revealed that while in the SAD group RTs on angry judgments were significantly faster relative to happy judgments, $t_{26}=2.86$, $p<0.05$, Cohen's $d=0.55$, no such difference was found between RTs for angry and happy judgments in the control group, $t_{20}=-0.77$, $p>0.40$, Cohen's $d=0.17$. Figure 2b displays group differences in RT bias scores reflecting mean RT for angry judgements minus mean RT for happy judgements. An additional set of follow-up contrasts revealed that RTs for happy judgments were significantly slower in the SAD group compared to the control group, $t_{46}=2.84$, $p<0.05$, Cohen's $d=0.83$, while no significant difference was found between the groups in RTs for angry judgments, $t_{46}=1.5$, $p>0.1$, Cohen's $d=0.44$. However, an additional

standardized RT analysis¹ suggested that the groups differed also in their reaction times for angry interpretations, which were faster in the SAD group compared to the control group. This suggests that the negative bias seen in the SAD group may be driven by a differential pattern in the judgment of positive and negative emotion.

¹ Since a significant main effect of group was found in the raw RT analysis, we conducted an additional analysis using standardized RTs. Standardized RTs were calculated as angry mean RT/total mean RT; and happy mean RT/total mean RT. Standardized RTs were submitted to a 2 x 2 ANOVA with Group (SAD, Control) as a between subjects factor and Judgment (happy, angry) as a repeated within-subject factor. As in the raw RT analysis, the main effect of Judgment was not significant, $F(1,46)=3.52$, $p>0.06$, $\eta^2_p=0.07$, and a main effect of group indicated that control participants were overall faster than participants with SAD, $F(1,46)=4.90$, $p<0.05$, $\eta^2_p=0.10$. This main effect was qualified by a significant Group-by-Judgment interaction effect, $F(1,46)=8.10$, $p<0.01$, $\eta^2_p=0.15$. Follow-up contrasts revealed that while in the SAD group standardized RTs to angry judgments were significantly faster relative to happy judgments, $t_{26}=3.18$, $p<0.01$, Cohen's $d=0.61$, no such difference was found between standardized RTs for angry and happy judgments in the control group, $t_{20}=0.79$, $p>0.40$, Cohen's $d=0.03$. An additional set of follow-up contrast revealed that standardized RTs for happy judgments were significantly slower in the SAD group compared to the control group, $t_{46}=2.77$, $p<0.01$, Cohen's $d=0.86$, and standardized RTs for angry judgments were significantly faster in the SAD group compared to the control group, $t_{46}=2.85$, $p<0.01$, Cohen's $d=0.83$.

4. Discussion

The results of the current study suggest that individuals with SAD show a negative interpretation bias when processing ambiguous social information. Specifically, we found this bias when participants were asked to label the emotion on faces containing varying degrees of conflicting angry/happy affect. Compared to non-anxious participants, individuals with SAD tended to make more negative (or less positive) interpretations of ambiguous faces. Also, participants with SAD tended to be slower when making positive (happy) interpretations compared to negative (angry) interpretations. This RT bias was not present among non-anxious participants. Moreover, a main effect of group was found, such that participants with SAD were overall slower when judging facial expressions. However, this main effect was subsumed under a group-by-emotion interaction, where follow-up contrasts demonstrated associations specifically with judgment of positive emotion (see Richards 2002 Study1 for a similar result). Taken together, this suggests that SAD involves a tendency to resolve ambiguity in facial expressions in a more negative (less positive) manner than non-anxious controls.

Similar findings have been found among abused children who, compared to non-abused children, tend to identify ambiguous faces as more angry on morphed continuums ranging from anger to sadness and anger to fear (Pollak and Kistler, 2002). It is possible that the cognitive mechanism underlying enhanced identification of angry faces could in fact be similar in these two populations. According to Pollak and Kistler (2002), abused children adapt to an environment where parental facial anger is an important predictor of potentially threatening, aversive interpersonal experience. A similar process may apply for socially anxious individuals, who experience others' anger or disapproval as signaling a threatening aversive experience

in social interaction. Thus they may become extremely sensitive or reactive to even mild expressions of anger.

Two previous studies have examined judgment of angry-happy morphed faces in clinical SAD samples (Garner et al., 2009; Jusyte and Schönenberg, 2014). Both studies found no differences in interpretation patterns of angry-happy morphed faces between SAD and control participants. Methodological differences may explain this discrepancy. In both previous studies, the presentation duration of the faces was longer compared to the current study, and on each trial participants rated the face on an emotional intensity scale, rather than only making a two-forced choice emotional judgment. These differences in presentation duration and response alternatives may have differentially engaged deliberative interpretation processes.

The current findings also support some now classic cognitive models of social anxiety (Clark, 2001; Clark and Wells, 1995; Rapee and Heimberg, 1997). These models suggest that socially anxious individuals possess negative assumptions or mental representations of both themselves and of others' evaluation of them. These assumptions lead individuals with SAD to appraise social situations as more threatening, to interpret benign or ambiguous cues as signs of disapproval or negative evaluation by others, and to interpret mildly negative social cues in an exaggerated negative manner. According to these models, individuals with SAD are highly sensitive to signs of impending negative evaluation, and differ from those without SAD in terms of the extent to which they appraise cues as predictive of threat. The current findings are in line with these predictions, as individuals with SAD tended to appraise the ambiguous faces as more negative compared with non-anxious controls. Moreover participants with SAD tended to make these negative appraisals of ambiguous faces faster than they did positive appraisals, again suggesting greater

sensitivity to ambiguous cues of social threat. Trower and Gilbert (1989) suggested that social anxiety has developed as a means of dealing with intra-species threat. According to this suggestion social anxiety is related to a defensive mechanism leading to a tendency to appraise others as hostile dominants and respond with submissive behavior in order to keep safe. The current findings that individuals with SAD tend to interpret more ambiguous faces as angry and to make angry interpretations faster than happy interpretations correspond with this theoretical framework as anger is the facial expression most typically seen as an "attack" expression, and is associated with traits of high social dominance and low affiliation (Knutson, 1996; Montepare and Dobish, 2003). Thus, interpreting facial expressions as angry may serve as a "safety mode" for socially anxious individuals under conditions of ambiguity and uncertainty, in order for them to avoid conflict with a potentially dominant and hostile person.

The current findings should be regarded in light of potential limitations and directions for future research: 1) In the current study we specifically focused on the recognition of anger as a negative expression, since anger may reflect others' disapproval and discontent, which is a primary concern for socially anxious individuals. Thus, we hypothesized that a tendency to interpret ambiguous faces as angry may be associated with SAD. However, this focus on only one type of negative emotion limits our ability to ascertain whether SAD patients demonstrate a specific tendency to interpret ambiguous faces as angry or, alternatively, a more general tendency to interpret ambiguous faces as negative. Future studies could explore this question by using stimuli consisting of various negative emotions (e.g., fear and disgust), or ambiguous emotions (e.g., surprise) and tasks that include other negative response options in addition to anger. This could help disambiguate whether this bias

363 is specific to anger or more general to any negative emotional expression. Also,
364 adding a neutral response option might differentiate whether the pattern of more
365 negative interpretation was due to more anger interpretations, less happy
366 interpretations, or both. However, whether the bias is specific toward anger
367 interpretation or toward negative interpretations in general, from a clinical point of
368 view, a CBM method which enhances positive interpretation of ambiguous faces
369 could be relevant. 2) The ecological validity of angry-happy morphed faces is not
370 entirely clear. While ambiguous facial expressions are common in everyday social
371 situations, the specific combination of angry and happy may be less common. Since
372 effective CBM protocols successfully applied angry-happy morphed faces (Penton-
373 Voak et al., 2013; Stoddard et al., 2016), we used similar stimuli here. Ecologic
374 validity aside, these stimuli do confront the participant with a forced choice of truly
375 ambiguous socially-related stimuli. Future studies may extend this stimuli selection
376 and use additional emotional combinations. 3) The ecological validity may be
377 somewhat limited also due to the use of only four different actors to generate the
378 morphed sequences. Using a small number of facial identities is not uncommon in
379 research using morphed faces (e.g., Gilboa-Schechtman et al., 2008; Reeb-Sutherland
380 et al., 2015; Richards et al., 2002) , and the faces for the current study were selected
381 so that they represent both genders and various ethnicities. However, future studies
382 may consider using a larger and more diverse set of actors in order to improve the
383 ecological validity and generalizability of this phenomenon. 4) The current design
384 does not allow definite conclusions on whether the two groups differ in their
385 sensitivity to emotional cues or in their response criterion to select angry or happy
386 judgments. The specific origin of the tendency to interpret ambiguous faces as angry
387 may be further examined in studies using paradigms that differentiate sensitivity and

criterion biases (e.g., Frenkel and Bar-Haim, 2011; Richards et al., 2002; Yoon et al., 2014). Moreover, to rule out the possibility of a more general negative response bias not related to interpersonal or emotional contexts, future studies could add trials requiring judgment of non-social stimuli such as objects. A previous study that applied this approach in the interpretation of verbal stimuli resulted in no anxiety-related between-group differences in interpretation of non-social scenarios (e.g., Voncken et al., 2003). Nonetheless, from a clinical perspective, whether this bias is due to sensitivity or due to response bias, it might be modifiable by CBM. 5) Based on the current sample, it is not possible to conclude whether the observed group differences are specific to SAD or, alternatively, characterize anxiety disorders in general. Future studies could include participants with other types of anxiety disorders (e.g., Generalized Anxiety Disorder). Finally, 6) the current findings indicate a large effect size for the between group difference in percent of angry interpretations. However, it is not possible to determine the clinical meaning of such a difference. The clinical relevance of this difference in SAD could be tested in future CBM studies measuring the impact of reduction in this bias on symptomatic behavior and mood.

In summary, the current findings provide evidence for a negative bias in interpreting ambiguous facial expressions in SAD, and lay the ground work for targeting this specific bias using established CBM techniques (Penton-Voak et al., 2013; Stoddard et al., 2016). To the best of our knowledge no CBM program has targeted interpretation biases related to facial expressions in SAD. The efficacy of such intervention could now be tested in future randomized controlled trials.

Funding information

412 This study was partially supported by US Department of Defense Grant
413 number W81XWH-13-2-0001. This sponsor was not involved and did not influence
414 study design, data collection, analysis, interpretation, writing or decision regarding
415 submission for publication.

416

417 **Table 1.** Frequencies, Means, and SDs of participants' characteristics by group.

418

	SAD	Control
Gender (F/M)	14/13	12/9
Age	28.4 (6.8)	25.5 (3.5)
LSAS Score	82.2 (12.4)	17.9 (7.7)
Accuracy	94.9 (5.0)	94.9 (6.1)
Raw RT Happy (ms)	739 (110)	649 (107)
Raw RT Angry (ms)	705 (118)	655 (107)

419

420 **Figure 1.** An example of a morphed sequence.

421

422 **Figure 2.** Percent of anger responses (A), and RT bias (B) for the SAD (solid bar)
423 and Control (striped bar) groups. Bars indicate standard errors.

References

- Amir, N., Foa, E.B., Coles, M.E., 1998a. Automatic activation and strategic avoidance of threat-relevant information in social phobia. *Journal of Abnormal Psychology* 107 (2), 285.
- Amir, N., Foa, E.B., Coles, M.E., 1998b. Negative interpretation bias in social phobia. *Behaviour Research and Therapy* 36 (10), 945-957.
- Amir, N., Taylor, C.T., 2012. Interpretation training in individuals with generalized social anxiety disorder: a randomized controlled trial. *Journal of Consulting and Clinical Psychology* 80 (3), 497.
- 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.
- Beard, C., Amir, N., 2009. Interpretation in social anxiety: when meaning precedes ambiguity. *Cognitive Therapy and Research* 33 (4), 406-415.
- Clark, D.M., 2001. A cognitive perspective on social phobia, in: Crozier, W.R., Alden, L.E. (Eds.), *International handbook of social anxiety*. Wiley, Chichester, UK, pp. 405–430.
- Clark, D.M., Wells, A., 1995. A cognitive model of social phobia. *Social phobia: Diagnosis, assessment, and treatment* 41 (68), 00022-00023.
- Coles, M.E., Heimberg, R.G., Schofield, C.A., 2008. Interpretation of facial expressions and social anxiety: Specificity and source of biases. *Cognition and Emotion* 22 (6), 1159-1173.
- Frenkel, T.I., Bar-Haim, Y., 2011. Neural activation during the processing of ambiguous fearful facial expressions: An ERP study in anxious and nonanxious individuals. *Biological Psychology* 88 (2-3), 188-195.

450 Fresco, D.M., Coles, M.E., Heimberg, R.G., Liebowitz, M.R., Hami, S., Stein, M.B.,
 451 Goetz, D., 2001. The Liebowitz Social Anxiety Scale: a comparison of the
 452 psychometric properties of self-report and clinician-administered formats.
 453 Psychological Medicine 31, 1025-1035.

454 Garner, M., Baldwin, D.S., Bradley, B.P., Mogg, K., 2009. Impaired identification of
 455 fearful faces in Generalised Social Phobia. Journal of affective disorders 115
 456 (3), 460-465.

457 Gilboa-Schechtman, E., Foa, E., Vaknin, Y., Marom, S., Hermesh, H., 2008.
 458 Interpersonal sensitivity and response bias in social phobia and depression:
 459 labeling emotional expressions. Cognitive Therapy and Research 32 (5), 605-
 460 618.

461 Gilboa-Schechtman, E., Foa, E.B., Amir, N., 1999. Attentional biases for facial
 462 expressions in social phobia: The face-in-the-crowd paradigm. Cognition &
 463 Emotion 13 (3), 305-318.

464 Gutiérrez-García, A., Calvo, M.G., 2014. Social anxiety and interpretation of
 465 ambiguous smiles. Anxiety, Stress & Coping 27 (1), 74-89.

466 Heuer, K., Lange, W.-G., Isaac, L., Rinck, M., Becker, E.S., 2010. Morphed
 467 emotional faces: Emotion detection and misinterpretation in social anxiety.
 468 Journal of Behavior Therapy and Experimental Psychiatry 41 (4), 418-425.

469 Huppert, J.D., Foa, E.B., Furr, J.M., Filip, J.C., Mathews, A., 2003. Interpretation bias
 470 in social anxiety: A dimensional perspective. Cognitive Therapy and Research
 471 27 (5), 569-577.

472 Huppert, J.D., Pasupuleti, R.V., Foa, E.B., Mathews, A., 2007. Interpretation biases in
 473 social anxiety: Response generation, response selection, and self-appraisals.
 474 Behaviour Research and Therapy 45 (7), 1505-1515.

475 Joormann, J., Gotlib, I.H., 2006. Is this happiness I see? Biases in the identification of
 476 emotional facial expressions in depression and social phobia. *Journal of*
 477 *Abnormal Psychology* 115 (4), 705-714.

478 Jusyte, A., Schönenberg, M., 2014. Threat processing in generalized social phobia:
 479 An investigation of interpretation biases in ambiguous facial affect. *Psychiatry*
 480 *Research* 217 (1), 100-106.

481 Knutson, B., 1996. Facial expressions of emotion influence interpersonal trait
 482 inferences. *Journal of Nonverbal behavior* 20 (3), 165-182.

483 Levin, J.B., Maron, S., Gur, S., Wechter, D., Hermesh, H., 2002. Psychometric
 484 properties and three proposed subscales of a self-report version of the
 485 Liebowitz Social Anxiety Scale translated into Hebrew. *Depression and*
 486 *Anxiety* 16, 143-151.

487 Liebowitz, M.R., 1987. Social phobia. *Modern problems of pharmacopsychiatry* 22,
 488 141-173.

489 Mathews, A., Mackintosh, B., 2000. Induced emotional interpretation bias and
 490 anxiety. *Journal of Abnormal Psychology* 109 (4), 602-615.

491 Mennin, D.S., Fresco, D.M., Heimberg, R.G., Schneier, F.R., Davies, S.O.,
 492 Liebowitz, M.R., 2002. Screening for social anxiety disorder in the clinical
 493 setting: using the Liebowitz Social Anxiety Scale. *Journal of Anxiety*
 494 *Disorders* 16 (6), 661-673.

495 Mogg, K., Philippot, P., Bradley, B.P., 2004. Selective attention to angry faces in
 496 clinical social phobia. *Journal of Abnormal Psychology* 113 (1), 160-165.

497 Montepare, J.M., Dobish, H., 2003. The contribution of emotion perceptions and their
 498 overgeneralizations to trait impressions. *Journal of Nonverbal behavior* 27 (4),
 499 237-254.

500 Muris, P., Merckelbach, H., Damsma, E., 2000. Threat perception bias in nonreferred,
 501 socially anxious children. *Journal of Clinical Child Psychology* 29 (3), 348-
 502 359.

503 Murphy, R., Hirsch, C.R., Mathews, A., Smith, K., Clark, D.M., 2007. Facilitating a
 504 benign interpretation bias in a high socially anxious population. *Behaviour*
 505 *Research and Therapy* 45 (7), 1517-1529.

506 Penton-Voak, I.S., Thomas, J., Gage, S.H., McMurrin, M., McDonald, S., Munafò,
 507 M.R., 2013. Increasing recognition of happiness in ambiguous facial
 508 expressions reduces anger and aggressive behavior. *Psychological Science* 24
 509 (5), 688-697.

510 Philippot, P., Douilliez, C., 2005. Social phobics do not misinterpret facial expression
 511 of emotion. *Behaviour Research and Therapy* 43 (5), 639-652.

512 Pollak, S.D., Kistler, D.J., 2002. Early experience is associated with the development
 513 of categorical representations for facial expressions of emotion. *Proceedings*
 514 *of the National Academy of Sciences of the United States of America* 99 (13),
 515 9072-9076.

516 Rapee, R.M., Heimberg, R.G., 1997. A cognitive-behavioral model of anxiety in
 517 social phobia. *Behaviour Research and Therapy* 35 (8), 741-756.

518 Reeb-Sutherland, B.C., Rankin Williams, L., Degnan, K.A., Pérez-Edgar, K.,
 519 Chronis-Tuscano, A., Leibenluft, E., Pine, D.S., Pollak, S.D., Fox, N.A., 2015.
 520 Identification of emotional facial expressions among behaviorally inhibited
 521 adolescents with lifetime anxiety disorders. *Cognition and Emotion* 29 (2),
 522 372-382.

523 Richards, A., French, C.C., Calder, A.J., Webb, B., Fox, R., Young, A.W., 2002.
524 Anxiety-related bias in the classification of emotionally ambiguous facial
525 expressions. *Emotion (Washington D C)* 2 (3), 273-287.

526 Rytwinski, N.K., Fresco, D.M., Heimberg, R.G., Coles, M.E., Liebowitz, M.R.,
527 Cissell, S., Stein, M.B., Hofmann, S.G., 2009. Screening for social anxiety
528 disorder with the self-report version of the Liebowitz Social Anxiety Scale.
529 *Depression and Anxiety* 26 (1), 34-38.

530 Sheehan, D.V., Lecrubier, Y., Sheehan, K.H., Amorim, P., Janavs, J., Weiller, E.,
531 Hergueta, T., Baker, R., Dunbar, G.C., 1998. The Mini-International
532 Neuropsychiatric Interview (MINI): the development and validation of a
533 structured diagnostic psychiatric interview for DSM-IV and ICD-10. *Journal*
534 *of Clinical Psychiatry*.

535 Stein, M.B., Goldin, P.R., Sareen, J., Zorrilla, L.T.E., Brown, G.G., 2002. Increased
536 amygdala activation to angry and contemptuous faces in generalized social
537 phobia. *Archives of general psychiatry* 59 (11), 1027-1034.

538 Stoddard, J., Sharif-Askary, B., Harkins, E.A., Frank, H.R., Brotman, M.A., Penton-
539 Voak, I.S., Maoz, K., Bar-Haim, Y., Munafo, M., Pine, D.S., 2016. An Open
540 Pilot Study of Training Hostile Interpretation Bias to Treat Disruptive Mood
541 Dysregulation Disorder. *Journal of Child and Adolescent*
542 *Psychopharmacology*.

543 Stopa, L., Clark, D.M., 2000. Social phobia and interpretation of social events.
544 *Behaviour Research and Therapy* 38 (3), 273-283.

545 Tottenham, N., Tanaka, J.W., Leon, A.C., McCarry, T., Nurse, M., Hare, T.A.,
546 Marcus, D.J., Westerlund, A., Casey, B.J., Nelson, C., 2009. The NimStim set

547 of facial expressions: Judgments from untrained research participants.
548 Psychiatry Research 168, 242-249.

549 Trower, P., Gilbert, P., 1989. New theoretical conceptions of social anxiety and social
550 phobia. Clinical Psychology Review 9 (1), 19-35.

551 Voncken, M.J., Bögels, S.M., de Vries, K., 2003. Interpretation and judgmental biases
552 in social phobia. Behaviour Research and Therapy 41 (12), 1481-1488.

553 Wilson, E.J., MacLeod, C., Mathews, A., Rutherford, E.M., 2006. The causal role of
554 interpretive bias in anxiety reactivity. Journal of Abnormal Psychology 115
555 (1), 103-111.

556 Yoon, K.L., Yang, J.-W., Chong, S.C., Oh, K.J., 2014. Perceptual Sensitivity and
557 Response Bias in Social Anxiety: An Application of Signal Detection Theory.
558 Cognitive Therapy and Research 38 (5), 551-558.

559

560

561

562 Figure 1

563
564

Figure 2