BRIEF REPORTS

Interpreting Neutral Faces as Threatening Is a Default Mode for Socially Anxious Individuals

K. Lira Yoon and Richard E. Zinbarg
Northwestern University

The authors of the present study used an incidental learning paradigm to investigate the interpretation of neutral facial expressions in socially anxious individuals. Participants were asked to detect the location of a target following the presentation of a facial picture (i.e., cue). Unbeknownst to participants, the target location was contingent on the valence of the cue, and participants thus learned to associate different target locations with either positive or negative facial expressions. The authors subsequently used this learned association to assess interpretive biases. If socially anxious individuals interpret neutral faces in a negative manner, they should be faster to detect a target that appears in the location that is associated with negative face cues when the target is presented after a neutral face cue. The authors also assessed whether the anticipation of a feared situation influenced interpretive biases by comparing participants with and without a speech threat on this task. Results indicate that socially anxious individuals are characterized by an interpretive bias regardless of the threat manipulation. In contrast, nonanxious individuals interpreted neutral faces in a negative manner only when they were in the threat condition.

Keywords: social anxiety; interpretive bias; face

As the basic fear in social anxiety (SA) is that of receiving negative evaluation, which is often conveyed by and inferred from facial expressions, interpretations of ambiguous facial expressions might be particularly relevant for understanding SA (Philippot & Douilliez, 2005). Evidence for a facial interpretive bias in socially anxious individuals, however, is weak at best. Although a few studies have reported that socially anxious individuals interpret neutral faces in a negative manner (e.g., Lundh & Öst, 1996a), others have failed to do so (e.g., Lundh & Öst, 1996b). Even more problematic, the studies reporting significant findings are mostly limited by their sole reliance on self-reports. It is, thus, possible that the results from previous studies might reflect socially anxious individuals' tendency to select negative response options (i.e., response bias).

We employed incidental learning methodology (e.g., Buchner & Wippich, 1998; Lewicki, 1986) to compare high-SA individuals with low-SA individuals in their interpretations of pictures of ambiguous facial expressions. As participants were not asked to direct their attention to the valence of the stimuli or to engage in conscious categorization of stimuli in terms of valence, we were able to study interpretive bias that could not be readily explained by response bias or demand.

The participants’ task was to locate a target and then to press a button corresponding to its location as soon as possible. Unbeknownst to the participants, there was a relationship between the valence of a face stimulus (i.e., cue) and the target location that participants were exposed to during the learning phase. In the learning phase, negative and positive faces served as cues. If learning occurred, participants should more rapidly detect the target appearing at the expected location according to the pattern presented in the learning phase. For the critical trials in the testing phase, neutral faces—inherently ambiguous as to the emotional state of the actor—appeared as cues.

It is important to note that threat manipulations often interact with individual differences in trait anxiety in regard to cognitive biases (e.g., Mathews & MacLeod, 1994). The results are inconsistent with some studies that have suggested threat manipulations enhance biases (e.g., Calvo & Castillo, 1997; Mansell, Clark, & Ehlers, 2003), whereas other studies have suggested that threat manipulations suppress biases in the anxious populations (e.g., Amir et al., 1996; Mathews & Sebastian, 1993). It thus seems important to examine the effects of threat manipulation on interpretive bias. Therefore, half the participants were assigned to a speech condition, and half were assigned to a no-speech condition.

On the basis of cognitive models of social phobia (Clark & Wells, 1995; Rapee & Heimberg, 1997), we hypothesized that socially anxious individuals would interpret neutral faces in a negative manner (i.e., they would show relative facilitation in target detection when the target appears at the location that they learned to associate with the negative faces following a neutral face cue).
Method

Participants

One hundred sixteen introductory psychology students completed the Social Phobia Scale (SPS; Mattick & Clarke, 1998) during an initial group testing session. Sixty students who scored in the top (SPS ≥ 18) or bottom (SPS ≤ 7) 25th percentile on the SPS were then chosen to participate in the study as part of a course requirement. Due to technical failure, 11 participants were unable to complete the incidental learning task. In addition, 2 participants had at least one missing cell and thus were excluded from analyses. The loss of data was spread evenly across SA level (i.e., 7 participants in the low-SA group and 6 in the high-SA group). The final sample consisted of 47 participants (35 female), with 23 (17 female) low-SA (SPS, M = 5.91, SD = 3.73) and 24 (18 female) high-SA participants (SPS, M = 28.75, SD = 8.36). About two thirds of our high-SA participants met or exceeded the score of 26 suggested by Peters (2000) to identify those with social anxiety disorder.

Materials

Symptom measures. The SPS assesses a participant’s typical levels of fear of scrutiny when performing a task or being observed by others and has adequate reliability and concurrent validity. Participants also completed the Beck Depression Inventory (BDI; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961), which has adequate reliability and has been extensively validated (e.g., Beck, Steer, & Garbin, 1988). Alpha in the current study equaled .93 for the SPS and .87 for the BDI. Participants’ state anxiety level was measured by a single item (“How anxious do you feel right now?”) on a 0 (not at all) to 9 (extremely) scale on the computer.

Face stimuli. Angry, disgusted, happy, and neutral faces were obtained from Ekman and Friesen’s (1976) Pictures of Facial Affect. Fourteen models (8 female, 6 male) posed for each facial affect, resulting in a total of 56 different facial stimuli. Eight neutral pictures appeared once, and the other 6 appeared twice; 6 happy pictures appeared eight times, and the other 8 appeared nine times; 10 angry pictures and 10 pictures showing disgust appeared six times; and the remaining 4 angry and 4 disgusted pictures appeared five times.

Incidental Learning Task

The task consisted of a learning phase and a testing phase. Each trial consisted of a 500-ms fixation point, a facial picture (cue) presented for 675 ms that was followed immediately by a target (letter L), and a question that asked the participant whether the person in the picture was a man or a woman to ensure attention to the stimuli.

During the learning phase, the cue was either negative (i.e., an angry or a disgusted face) or positive (i.e., a happy face). Once the cue disappeared, the target appeared at any of four premarked positions on the computer screen (see Figure 1). When the cue was negative, the target appeared at either the top or the bottom two locations (i.e., negative location) on 80% of the trials. The side opposite to the negative location was the positive location. For example, for some participants when the cue was negative, the target appeared at the upper-left location 40% of the time, at the upper-right location 40% of the time, and at each of the two lower locations 10% of the time. For these participants, when the cue was positive, the target appeared at the bottom-left 40% of the time and at the bottom-right 40% of the time. Whether the top or the bottom was designated as the negative location was determined randomly for each of the participants. For the final sample, there was no difference in the number of participants from each group who received the top as the negative location.

Participants were instructed to respond to the appearance of the target as quickly as possible by pressing one of four keys that corresponded to the locations of the target. We measured response latency as the time interval between the onset of the target display and the response. Each participant was presented with 10 blocks of learning trials, with each block consisting of 20 trials (i.e., 10 trials with negative cues—five angry and five disgusted—and 10 trials with positive cues) in a different randomized order.

During the testing phase, neutral face cues appeared in addition to positive and negative face cues. The relationship between the valence of the cue and the target location during the testing phase remained identical for the positive and the negative cues. For the neutral cue trials, the target appeared at any of the four possible positions with equal chances. There were 20 trials each with neutral, positive, and negative cues. We used the same positive and negative face cues as we did in the learning phase.

Procedure

Participants were tested in groups of three to four. After signing the consent form, participants in the speech condition were asked to prepare for a speech for 2 min, whereas participants in the no-speech condition were asked to write an essay. Participants in the no-speech condition were assured that their essays would not

![Figure 1](https://example.com/figure1.png)
be read by anyone but that the experimenter would check whether they wrote an essay. Participants in the speech condition were told that one of them would be randomly selected to give a videotaped speech toward the end of the session. Next, participants were asked to complete the computer task. Participants were told that their task was to identify the location of the L by pressing one of the four keys that corresponded to the location of the target. After responding to the target, participants indicated whether the person in the picture was a man or a woman.

Participants began the incidental learning task by completing the single-item state anxiety measure. One of the participants in each speech group was asked to give a 2-min videotaped speech once everyone in the group completed the incidental learning task. After the speech (or after the computer task, if they were in the no-speech group), participants were given a questionnaire that assessed awareness of the contingencies.

Results

Participant Characteristics

A series of SA × Speech analyses on age, gender, BDI, and SPS scores were conducted. Socially anxious individuals scored higher on the BDI, \( F(1, 43) = 12.16, p < .001 \), and on the SPS, \( F(1, 43) = 139.46, p < .001 \), than their counterparts. No other effects were significant, indicating that randomization was successful.

Manipulation Check

To test whether the speech manipulation had its intended effect, we examined participants’ anxiety levels right before they began the computer task (i.e., after they learned their experimental condition). The SA × Speech analysis of variance (ANOVA)\(^1\) revealed significant main effects for SA, \( F(1, 43) = 12.08, p < .001 \), \( \eta^2 = .05 \), and for speech, \( F(1, 43) = 4.30, p < .05, \eta^2 = .02 \). The pattern confirmed that groups differed in the manner that we intended by the manipulation (speech, \( M = 3.83, SD = 1.81 \); no-speech, \( M = 2.87, SD = 1.87 \)). Importantly, the interaction between SA and speech was not significant, \( F(1, 43) = 2.06, \eta^2 = .007 \). Means and standard deviations of state anxiety level for each group was as follows: low-SA, no-speech (\( M = 2.36, SD = 1.86 \)); low-SA, speech (\( M = 2.67, SD = 1.30 \)); high-SA, no-speech (\( M = 3.33, SD = 1.83 \)); high-SA, speech (\( M = 5.00, SD = 1.48 \)).

Data Reduction

Data from trials with errors (i.e., incorrect identification of the target location or incorrect identification of the gender of a cue) were discarded. In addition, response times (RTs) that were less than 200 ms or greater than 1,500 ms were excluded on the basis of the boxplot (cf. Barnett & Lewis, 1994). For the final sample, the mean percentages of data lost due to errors and outliers were 2.9% and 1.5%, respectively. Errors and outliers did not differ by group, speech condition, or cue type (all \( F_s < 1 \)). The overall mean RT was 609.1 ms (\( SD = 451.4 \)) before and 563.9 ms (\( SD = 122.5 \)) after the data cleanup. Because our RT data were not normally distributed (Shapiro–Wilk = .89, \( p < .001 \)), we reanalyzed the data with log-transformed RTs, which successfully normalized the data (Shapiro–Wilk = .96, \( n.s. \)). The results remained the same, therefore we report the analyses based on raw RTs for the simplicity of interpretation of results.

Learning of the Cue Valence and Target Location Association

To examine whether participants learned the relationship between cue valence and the target location and to rule out the possibility that socially anxious individuals learned the association involving the negative cues more strongly than did their less anxious counterparts, we carried out an SA (high vs. low) × Cue Type (negative vs. positive) × Target Location (expected vs. unexpected) × Block mixed ANOVA with polynomial trend analysis on the RTs for the training trials.\(^2\) As expected, the Target Location × Linear Trend of Block interaction was significant, \( F(1, 45) = 4.53, p < .04, \eta^2 = .01 \); the decrease in RT was steeper over blocks for the expected target location (slope = −.007) compared with the unexpected target location (slope = −.003). The Cue Type × Target Location × Linear Trend of Block interaction was not significant (\( F < 1 \)). Importantly, the four-way interaction, which might indicate differential acquisition of the associations between the groups, was not significant (\( F < 1 \)). In addition, neither the SA × Target Location × Block interaction nor the SA × Target Location was significant (both \( F_s < 1 \)). Therefore, these results provide no support for a stronger expectation for the threat location among the socially anxious participants.

Interpretive Bias and SA

Although the results above suggest that participants learned the contingencies to the same extent regardless of their SA level, it is still possible for socially anxious individuals to become faster at detecting the target that appears at the negative location regardless of the valence of the cue on the testing trials. Similarly, participants in the speech condition might have become better at detecting the target that appeared at the negative location due to their heightened levels of anxiety. To rule out these possibilities and to test the main hypothesis that socially anxious individuals interpret neutral faces negatively, a SA × Speech × Cue Type × Target Location ANOVA was conducted on the RTs during the testing phase. There was a significant main effect of target location, \( F(1, 43) = 7.65, p < .01, \eta^2 = .15 \), which was qualified by a significant four-way interaction, \( F(2, 42) = 3.69, p < .04, \eta^2 = .15 \). The interaction between SA and target location, however, was not significant, suggesting that socially anxious individuals were not significantly different from their less-anxious counterparts in their readiness to detect targets appearing at locations associated with negative cues. Likewise, the participants in the speech condition were not significantly different from those in the no-speech condition in their tendency to detect targets that appear at locations associated with negative cues as indicated by a nonsignificant Speech × Target Location interaction.

---

\(^1\) Because the size of a group that a participant was in could affect the level of state anxiety, an additional ANOVA with SA, speech condition, and group size was conducted. The results remained the same, and none of the effects involving the group size was significant.

\(^2\) Mean RTs for each block are available upon request from K. Lira Yoon.
To follow up on the significant four-way interaction, we tested the simple three-way interaction between SA, speech, and target location separately for neutral, positive, and negative cues. The simple three-way interaction was significant for neutral cues, $F(1, 43) = 10.56, p < .01, \eta^2 = .17$, but not for positive or negative cues (both $Fs < 1$), indicating that the group differences in responses to the neutral cues were responsible for the four-way interaction. Table 1 presents the means in each condition.

To break down the simple three-way interaction for neutral cues, we calculated a speeding index by subtracting the RT when the target appeared at the negative location. Thus, negative speeding indexes reflect a tendency to respond more rapidly when the target appeared at the negative location. When the speeding indexes for the neutral cues were subtracted from the RTs of detected targets that appeared at positive locations, the speeding index for the neutral cues was significantly lower than it was for positive cues in the no-speech, high-SA group, $t(11) = 2.78, p < .05$. No other comparisons were significant.

### Discussion

Through our use of an incidental learning paradigm, the results suggest an interpretive bias in socially anxious individuals that cannot be readily explained by either response bias or experimenter demand. Thus, the present results suggest that highly socially anxious individuals treat neutral social interaction cues as conveying anger or disgust and/or contempt, thereby extending the results of earlier studies in which researchers used verbal stimuli (e.g., Hirsch & Mathews, 2000). These findings also replicate and extend results obtained in a previous study in which researchers used a priming paradigm (Yoon & Zinbarg, 2007). Although demand characteristics cannot be conclusively refuted, it seems implausible that the participants would have figured out (a) our hypothesis that SA would be related to interpretive bias and (b) how this hypothesis translated into predictions about their RTs.

Another purpose of the current study was to examine whether the interpretive bias would be affected by a threat manipulation. In the current study, the group differences were significantly smaller when there was an explicit threat (i.e., speech condition). In fact, the group differences were only significant when there was no apparent threat, such that high-SA individuals exhibited a negative interpretive bias, whereas the low-SA individuals did not exhibit such a bias. In contrast, participants, regardless of anxiety levels, exhibited negative interpretive biases in the threat condition. Thus, it could be argued that interpreting neutral faces in a negative manner is the default mode for high-SA individuals, whereas low-SA individuals interpret neutral faces negatively only when anticipating threat.

One explanation for the effect of the speech threat draws on the model proposed by Mathews, Mackintosh, and Fulcher (1997).

---

**Table 1**

<table>
<thead>
<tr>
<th>Target location</th>
<th>Low-SA group</th>
<th></th>
<th>High-SA group</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Speech ($n = 12$)</td>
<td>No-speech ($n = 11$)</td>
<td>Speech ($n = 12$)</td>
<td>No-speech ($n = 12$)</td>
</tr>
<tr>
<td>Neutral cues</td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
<td>$SD$</td>
</tr>
<tr>
<td>Negative</td>
<td>467.71</td>
<td>57.77</td>
<td>593.16</td>
<td>201.21</td>
</tr>
<tr>
<td>Positive</td>
<td>514.13</td>
<td>96.38</td>
<td>553.55</td>
<td>169.24</td>
</tr>
<tr>
<td>Negative cues</td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
<td>$SD$</td>
</tr>
<tr>
<td>Negative</td>
<td>501.66</td>
<td>140.77</td>
<td>585.64</td>
<td>186.26</td>
</tr>
<tr>
<td>Positive</td>
<td>557.49</td>
<td>143.85</td>
<td>636.69</td>
<td>297.85</td>
</tr>
<tr>
<td>Positive cues</td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
<td>$SD$</td>
</tr>
<tr>
<td>Negative</td>
<td>549.17</td>
<td>240.51</td>
<td>625.68</td>
<td>212.66</td>
</tr>
<tr>
<td>Positive</td>
<td>507.23</td>
<td>95.60</td>
<td>594.61</td>
<td>163.71</td>
</tr>
</tbody>
</table>

**Note.** SA = social anxiety.
According to Mathews et al., interpretive biases arise from selective attention to threat, which enhances the activation of the threatening interpretation of an ambiguous stimulus. More specifically, we assume that when faced with the threat of public speaking, most people experience enhanced selective attention to potentially threatening faces. For the low-SA group, this resulted in a boost of the activation of the negative interpretation of an ambiguous face such that it tended to win the mutually inhibitory competition among the activated interpretations. For the high-SA group, the activation of the negative interpretation was already boosted to the point at which the negative interpretation was already tending to win even in the absence of the speech threat.

What remains unclear are the conditions in which a threat manipulation (a) enhances group differences in bias (e.g., Calvo & Castillo, 1997), (b) masks group differences by causing the low-SA group to demonstrate the bias otherwise characteristic of the high-SA group (as presented here), or (c) masks group differences by suppressing the bias otherwise demonstrated by the high-SA group (e.g., Mathews & Sebastian, 1993). Studies differ in several ways, including the type of bias assessed (i.e., attentional vs. interpretive), the type of stimuli presented (e.g., verbal, pictorial), the nature of the populations studied (i.e., clinical vs. high-trait anxious), and the temporal location of a threat. Further research needs to be done to sort out which methodological differences account for the differences in the results.

What remains unclear are the conditions in which a threat manipulation (a) enhances group differences in bias (e.g., Calvo & Castillo, 1997), (b) masks group differences by causing the low-SA group to demonstrate the bias otherwise characteristic of the high-SA group (as presented here), or (c) masks group differences by suppressing the bias otherwise demonstrated by the high-SA group (e.g., Mathews & Sebastian, 1993). Studies differ in several ways, including the type of bias assessed (i.e., attentional vs. interpretive), the type of stimuli presented (e.g., verbal, pictorial), the nature of the populations studied (i.e., clinical vs. high-trait anxious), and the temporal location of a threat. Further research needs to be done to sort out which methodological differences account for the differences in the results.

Limitations of this study are worth mentioning. First, we did not administer diagnostic interviews. Thus, future studies are necessary to test whether the obtained results can be generalized to a clinical population. Second, we do not have information about whether participants in the speech condition believed that they had to give a speech. However, considering the higher levels of anxiety in the speech condition compared with the no-speech condition, the speech manipulation seemed to have its intended effect. It is also possible that our low-SA group might have included individuals with high public speaking anxiety who are otherwise not socially anxious. This could have contributed to the findings of low-anxiety individuals in the speech condition who showed negative biases. Third, we used a single-item measure to assess the level of state anxiety. The item has good face validity and demonstrated that the speech manipulation had its intended effects on the level of state anxiety. It is still possible, however, that the single item might not have adequately assessed the group differences due to the threat manipulation. Relatedly, our results might represent a levels effect; that is, an increase from low to moderate state anxiety may serve to elicit a negative interpretive bias, which is unaffected by further increases in state anxiety. Despite the nonsignificant interaction between the group and the speech condition, the pattern of means suggests that our threat manipulation moved individuals with different levels of SA through different parts of the state anxiety continuum. Thus, it remains for future researchers to determine whether the difference between the anxiety groups that are not undergoing a social challenge in interpretive response is due to the inherent state anxiety difference between the groups or to a different impact of state anxiety on interpretive bias in the two groups.

One might argue that neutral faces are not ambiguous. We note that one can draw different inferences about the actor’s emotional state from a neutral expression, because there is considerable ambiguity in this process when the actor displays a neutral expression. Relatedly, Ekman and Friesen’s (1976) norms indicate that the neutral faces tend to be more ambiguous than other facial expressions, as there is far more variance in responses to the
neutral faces. It is possible that the present results might reflect socially anxious individuals’ negative emotional reactions to faces that they and the low-SA group interpret identically. Differential emotional appraisal alone, however, cannot explain the whole story. If the groups are making different emotional appraisals of the neutral faces, and socially anxious individuals find all faces more aversive than their less-anxious counterparts, we are still left with a continuum of relative aversiveness within each group. That is, differential emotional appraisal cannot fully account for why the socially anxious individuals tend to respond to neutral faces more as they do the truly aversive faces than they do the happy faces. Another important extension for future studies would be to use nonsocial stimuli in the learning phase to even more definitively rule out that the current results are due to group differences in initially learning about the positive and negative social stimuli.

These limitations notwithstanding, the incidental learning paradigm renders response bias implausible, as participants were never asked to indicate what expression they thought was represented in each of the pictures. Thus, the present study provides strong support for the hypothesis that SA is associated with a bias toward interpreting neutral facial expressions in a negative manner. More specifically, our results suggest that interpreting neutral faces in a negative manner is the default mode for socially anxious individuals, whereas people with low levels of anxiety interpret neutral faces negatively only when they are under threat.

References