



Reliability of an Interpretation Bias Task of Ambiguous Faces and Its Relationship with Social Anxiety, Depression, and Looming Maladaptive Style

Ángel Prieto-Fidalgo¹ · Sven C. Mueller^{1,2} · Esther Calvete¹

Accepted: 24 November 2022 / Published online: 6 December 2022
© The Author(s) 2022

Abstract

Social anxiety (SA) and depression have been associated with negative interpretation biases of social stimuli. Studies often assess these biases with ambiguous faces, as people with SA and depression tend to interpret such faces negatively. However, the test–retest reliability of this type of task is unknown. Our objectives were to develop a new interpretation bias task with ambiguous faces and analyse its properties in terms of test–retest reliability and in relation to SA, depression, and looming maladaptive style (LMS). Eight hundred sixty-four participants completed a task in which they had to interpret morphed faces as negative or positive on a continuum between happy and angry facial expressions. In addition, they filled out scales on SA, depressive symptoms, and LMS. Eighty-four participants completed the task again after 1–2 months. The test–retest reliability was moderate ($r = .57–.69$). The data revealed a significant tendency to interpret faces as negative for people with higher SA and depressive symptoms and with higher LMS. Longer response times to interpret the happy faces were positively associated with a higher level of depressive symptoms. The reliability of the present task was moderate. The results highlight associations between the bias interpretation task and SA, depression, and LMS.

Keywords Negative interpretation biases · Social anxiety · Depression · Looming maladaptive style

Cognitive models of social anxiety (SA) and depression propose a hierarchical model whereby maladaptive cognitive schemas guide attentional processes and facilitate biased interpretations congruent with these schemas (Amir et al., 1998;

✉ Ángel Prieto-Fidalgo
a.prieto@deusto.es

¹ Faculty of Health Sciences, University of Deusto, Avenida de Las Universidades, 24, 48007 Bilbao, Spain

² Department of Experimental Clinical and Health Psychology, Ghent University, Ghent, Belgium

Beck, 1976; Coles et al., 2008; Heimberg et al., 2014; Schultz & Heimberg, 2008). Specifically, in SA, models have hypothesized that interpretation biases are not only associated with SA, but also are a relevant maintenance mechanism (Clark & Wells, 1995; Heimberg et al., 2014). Similarly, it has been hypothesized that individuals with depression tend to create more negative interpretations of ambiguous information (Clark et al., 2000). Some authors consider that this negative interpretation of ambiguous information can be the cause of depression (Beck & Haigh, 2014).

Interpretation biases in ambiguous social scenarios have been examined mainly through two different lines of research. The first line includes studies that evaluate interpretation biases in imagined ambiguous social scenarios, which are verbally described to the participants (Dineen & Hadwin, 2004; Everaert et al., 2020; Sanchez et al., 2015). Commonly, these studies show a higher negative interpretation bias in participants with SA or depression. In fact, a meta-analysis has confirmed the association between the negative interpretation of ambiguous scenarios and verbal stimuli with SA ($g=0.97$; Chen et al., 2020). Another meta-analysis that included mainly studies employing this type of stimuli also found an overall association between negative interpretation and symptomatology of depression ($g=0.72$; Everaert et al., 2020).

The second line of research includes studies based on the interpretation of facial expressions. Faces are a crucial source of information to recognize the emotional state of others (Ekman & Friesen, 2003). Consequently, the information processing of ambiguous facial expressions is a fundamental aspect of social interactions. When these situations are interpreted negatively, hostile and critical intentions are attributed to others, which may contribute to triggering anxious and/or depressive feelings (Maoz et al., 2016). The method employed to assess interpretation bias using faces typically involves generated stimuli through intermediate (i.e., morphed) images between two facial expressions, each one with a different emotion, resulting in a battery of expressions of different intensity. For example, Jusyte and Schönberg (2014) combined fearful, happy, and angry faces and generated three sets of stimuli (happy–fearful, happy–angry, fearful–angry).

In terms of anxiety, some studies concluded that people with SA tend to classify ambiguous faces as less trustworthy (Gutiérrez-García & Calvo, 2016; Gutiérrez-García et al., 2019) and interpret ambiguous faces (Maoz et al., 2016) or even neutral or ambiguous faces as negative to a higher extent (Lira Yoon & Zinbarg, 2007; Prieto-Fidalgo et al., 2022). However, other studies did not find this tendency (Jusyte & Schönberg, 2014; Schofield et al., 2007). Moreover, the relationship between the interpretation of ambiguous faces and SA is not entirely clear. In one study, individuals with SA were sensitive to the identification of the expression of “fear” when comparing fear–anger or fear–happy faces, but no differences emerged when comparing happy–angry images (Garner et al., 2009). Another study with a similar methodology and stimulus set failed to replicate differences between the SA group and healthy participants (Jusyte & Schönberg, 2014). Similarly, no differences were found in the identification of facial morphs with happy and disgusted expressions as anchor points (Schofield et al., 2007). Despite the diversity in findings, which may be explained by the diversity of methodologies employed, a recent meta-analysis found that people with higher levels of SA tend to interpret images of faces more negatively with a medium effect size ($g=0.60$; Chen et al.,

2020). Some studies also encountered differences in the reaction time (RT) that needed to make a decision. For example, a study demonstrated that the social anxiety group was faster interpreting images as angry than as happy and was slower when making happy judgments relative to the control group (Maoz et al., 2016).

Regarding depression, a meta-analysis found that people with high depressive symptomatology made more negative interpretations as well as fewer positive interpretations in studies that included the evaluation of ambiguous verbal or visual stimuli (Everaert et al., 2017). However, only three of the 87 included studies used a face interpretation task (Beevers et al., 2009; Lee et al., 2016). One of the studies found a higher number of negative interpretations in ambiguous faces in a combination of happiness–sadness and happiness–fear in participants with depressive symptoms (Beevers et al., 2009). In addition, patients with major depression had greater general difficulty in detecting happy emotions when compared with neutral emotions (Soto et al., 2021). Additionally, a study has also found that people with depression require less time to correctly recognize faces of sadness, fear, and anger (Lee et al., 2016). Another study reported that the depressed group, compared with the control group, required more time to interpret happy faces (Leppänen et al., 2004). Shiroma and colleagues (2016) found that participants without depression were faster at identifying happy faces when compared with people with depression, but the differences were not statistically significant. In summary, studies suggest that people with depression require more time to process faces associated with positive emotions (happy faces) and require less time to process faces associated with negative emotions (sadness, fear, anger).

Indeed, cognitive mechanisms are assumed to be hierarchically organized, so interpretation biases would be guided by cognitive styles (Beck & Dozois, 2011; Clark & McManus, 2002). One example is the looming maladaptive style (LMS; Riskind et al., 2000). LMS is proposed as a maladaptive cognitive style, mainly associated with anxiety. It has been described as a danger schema that “produces schematic biases in the selection, interpretation, and recall of potential threat” (Riskind et al., 2000, p. 838). People with LMS would tend to perceive that a threat is going to get progressively worse. Typically, LMS has been associated with anxiety disorders (Adler & Strunk, 2010), including SA (Brown & Stopa, 2008), but not with depression (González-Díez et al., 2014; Riskind & Williams, 2005) and it has been demonstrated that it predicts maladaptive cognitions such as automatic thoughts (Calvete et al., 2016). In fact, it has been hypothesized that the LMS could influence interpretation biases (Riskind et al., 2000). For example, Riskind et al. (2000) asked participants to listen to homophone words (e.g., “die” vs “dye”) and found that LMS predicted the tendency to hear a greater number of threatening words (e.g., “die”). In addition, they also found a similar effect after the presentation of images, i.e., LMS was associated with the recall of a greater number of threatening pictures (explicit memory). Finally, LMS in that study was associated with a greater number of writing down of threatening words (implicit memory) after participants were asked to write down the first word that came to mind. In the same way, since it has been proposed that LMS would lead to a more biased interpretation and that faces are a relevant social cue (Ekman & Friesen, 2003), it can be expected that LMS related to social events would also be associated with a negative interpretation of ambiguous faces.

The disparity in the results within SA and between SA and depression may be due to the differences in methodologies. For example, while some studies are measuring responses dimensionally, considering the intensity of emotion (Garner et al., 2009; Jusyte & Schönberg, 2014) or the degree of confidence that it generates (Gutiérrez-García & Calvo, 2016), others could be requiring greater reflection on the part of the participants compared to more direct methodologies, such as a forced choice between two emotions (Maoz et al., 2016). Moreover, studies that force discrimination of ambiguous faces between two emotions (e.g., happy–angry) presuppose a negative valence for certain emotions (e.g., anger, sadness, fear, disgust) and a positive one for others (e.g., happy, being content). However, a happy face can be considered false or unreliable (Gutiérrez-García & Calvo, 2016). A final example of methodological differences lies within the stimuli themselves. For instance, using stimuli that include facial features highly associated with a specific emotion (e.g., faces with the mouth opened and visible teeth identified as smiling faces and faces with the mouth closed as angry faces; Jusyte & Schönberg, 2014) probably can lead to automatic responses that could prevent the activation of higher-order cognitive mechanisms associated with interpretation biases.

As previously discussed, theoretical models of social anxiety and depression consider interpretation biases as a relevant maintenance factor (Beck & Haigh, 2014; Clark & Wells, 1995; Clark et al., 2000; Heimberg et al., 2014). In addition, it is believed that tasks employing visual stimuli of faces may be more ecological than other types of psychometric tools for assessing interpretation biases (Heuer et al., 2010). Despite all this, to the best of our knowledge, no previous study on interpretation bias that uses images of facial expressions has examined the test–retest reliability of the instrument. Taking into account that reliability of a test is a significant component of its validity (Kappenman et al., 2014) and is necessary to establish the consistency of individual patterns (MacLeod et al., 2019), analyzing the reliability of this type of instrument is mandatory. For example, having a consistent instrument across time allows us to carry out longitudinal studies or analyze if interventions can modify interpretation bias. This is also a probe of the stability of the construct across time. The scarcity of interest in test–retest reliability in interpretation biases contrasts with in-depth analysis of the stability of tasks tapping into low level levels of processing—that do not require participants’ interpretation or identification—have been performed on attentional bias tasks (e.g., the dot probe) (Bantin et al., 2016). The results have indicated low levels of test–retest reliability (MacLeod et al., 2019), reducing enthusiasm for using this task. However, the test–retest reliability of interpretation bias tasks, which require higher level cognitive processes (identification or interpretation), remains to be determined.

Therefore, this study developed a facial-emotion task to assess interpretation bias based on some methodological improvements relative to prior work. First, as mentioned above, expressions of happiness may not be interpreted necessarily as positive. Therefore, methods that ask the participant to recognize faces as a concrete emotion can fail in identifying interpretation bias when a participant correctly recognizes a happy face as happy but interpret the face in a negative way (e.g., unreliable or untrustworthy). Thus, employing a distinct method from the classification of the emotion could better capture the interpretation bias (Gutiérrez-García et al.,

2016). For example, a forced choice between “positive” and “negative” options could be an adequate option. Second, to avoid an automatic response to signals that are very associated with a specific emotion (e.g., faces with an open mouth with happy emotion), following the recommendation of Jusyte and Schönenberg (2014), stimuli could be selected without this type of obvious signal (facial expressions with a closed mouth).

A primary purpose of study is to explore the functioning of the task, analyzing the test–retest reliability of the different indices of the task and analyzing sources of the validity of the task (relationship with SA, depression, and LMS). The secondary purpose of the study was to analyze the relationship between interpretation bias measured by means of the new interpretation bias task with ambiguous faces and symptoms of SA and depression, and LMS. Concerning primary purpose, we hypothesize that (1) the angrier the faces are displayed, the higher the rate of negative responses will be; (2) test-retest correlation coefficients will be similar to those reported by other studies that use the dot probe—the conceptually and methodologically most similar task that has previously assessed test–retest reliability—and (3) that the indexes of the task will be associated with SA, depression, and LMS, and these will discriminate between low and high SA and depression symptoms group. In relation to the secondary purpose, it was hypothesized that people with higher symptoms of SA and depression will label more ambiguous faces as negative, would need more time to label happy faces as positive, and would need less time to label anger faces as negative. The same tendency toward negative evaluation, slower response to label happy faces as positive, and faster response to label anger faces as negative would also be associated with the social LMS, but—according to the literature (Riskind & Williams, 2005)—not with the physical one.

Method

Design and Participants

We carried out a two-wave longitudinal study, including 864 first-year college or vocational training students from 10 different centers of Bizkaia (Spain). We calculated a posteriori power with G*Power (Faul et al., 2009) and indicated that which the number of participants and a power of 80%, it would be able to detect correlations of $r=0.10$ for cross-sectional analysis and of $r=0.30$ for test–retest correlations. The age of the participants ranged from 15 to 29 years (mean age = 19.54, $SD=2.51$, 44.4% women). In addition, 84 participants completed the measure again 1–2 months later in order to evaluate the test–retest reliability (mean age = 19.51, $SD=0.13$, 81.9% women).

Instruments

The Spanish version (Olivares et al., 2005) of the Social Anxiety Scale for Adolescents (SAS-A, La Greca & Lopez, 1998) was used to assess SA. The SAS-A is

composed of 18 items (e.g., “I am ashamed to be surrounded by people I do not know”) with three subscales: Fear of Negative Evaluation (FNE), Social Avoidance and Distress in New Situations or with Strangers (SAD-New), and Social Avoidance and Distress in General Situations or with People you Know (SAD-G). Each statement is rated on a five-point frequency-type scale ranging from 1 (*never*) to 5 (*all the time*). Studies have confirmed the internal consistency of the items and the three-factor structure of the SAS-A in a Spanish sample (Olivares et al., 2005). The overall internal consistency of the SAS-A was 0.93 (Cronbach’s alpha), with respective subscale coefficients of 0.89, 0.81, and 0.83 for the FNE, SAD-New, and SAD-G.

To measure depressive symptoms, the Center for Epidemiologic Studies-Depression (CES-D; Radloff, 1977, Spanish version: Calvete & Cardeñoso, 1999) was used. This is a 20-item questionnaire (e.g., “My appetite was poor”) rated on a four-point frequency scale from 0 (*rarely*) to 3 (*all the time*). Studies that have analyzed psychometrical properties of the CES-D in a Spanish sample have been reported excellent internal consistency, good sensitivity (77.1%), and good specificity (79.4%) (Ruiz-Grosso et al., 2012). In this study, Cronbach’s alpha coefficient was 0.92.

The Spanish version of the Looming Maladaptive Style Questionnaire (LMSQ-R; González-Díez et al., 2014; Riskind et al., 2000) was used to assess LMS. The questionnaire describes six potentially stressful scenarios (three physical and three social) and measures the tendency to estimate whether the risk of a potential threat will increase as well as expectations regarding their deterioration over time. In order to create a shorter battery of questionnaires, in this study, we used two physical scenarios (e.g., heart attack) and two social ones (e.g., social meeting) to assess both physical and social looming styles. The participant had to imagine each situation and rate it on a five-point Likert scale, with three questions related to the expectation of the threat. Studies have confirmed two second-order factors (social looming and physical looming) and measurement invariance of the test across gender in Spanish sample (González-Díez et al., 2014). In this study, the internal consistency for LMSQ-R measured by Cronbach’s alpha coefficient was 0.86 for the total score, 0.82 for Social Looming, and 0.80 Physical Looming.

Ambiguous Faces Interpretation Task

To develop a task for the evaluation of cognitive biases in the interpretation of ambiguous faces, a total of eight models (4 women and 4 men) were chosen from the Chicago Face Database (Ma et al., 2015). For each model, we selected three images—one classified as happy, one as neutral, and one as angry—following three criteria: (1) the model was not wearing any object such as glasses or a hair clip, (2) the model’s mouth was closed, and (3) the specifications of each image (framing, body postures of the models) allowed a clean morphing process. We used Morpheus Photo Morpher® v3.17 software to generate morphed faces. A set of ambiguous faces was generated by combining a happy face of each model with the respective neutral face of the same model and the other set by combining an angry face

with the respective neutral face of the same model. As a result, a set of nine levels of morphed faces—including the actual happy, angry, and neutral faces—was obtained, as shown in Fig. 1. Ultimately, a 72-face stimulus battery was generated (nine per model).

The task was developed in JavaScript, following recommendations for stimulus presentation (Garaizar & Reips, 2019). The presentation of each of the 72 faces was preceded by an orange fixation cross for 500 ms on a white background. Subsequently, each face was presented in the center of the screen (640 px wide and 448 px high) only once in random order until the response of the participant on a white background. After, a white screen was presented during 500 ms.

While faces were on the screen, the participants were asked to indicate whether the face presented showed a negative emotion by pressing the N key on the keyboard or P for the positive one. The participants were instructed to respond as quickly as possible to the task. The task recorded the responses and RT for each trial (for an illustration, see Fig. 2). Prior to the task, all participants completed four training trials with images not used in the task.

Procedure

The participants answered the scales in a fixed order—sociodemographics, SAS-A, LMSQ-R, CES-D, and ambiguous faces interpretation task—using a computer after providing informed consent. Then, they performed the interpretation bias task. The batteries of questionnaires and the task were presented in the Qualtrics® platform. After 1–2 months, 120 participants were contacted again to perform the interpretation bias task in-person in the laboratory, and 84 people did so. There were no significant mean differences between the participants that chose to answer in the laboratory and those that did not in SA, $t(118)=0.23$, $p=0.82$, symptoms of depression, $t(118)=0.34$, $p=0.73$, social LMS, $t(118)=1.01$, $p=0.32$, and total score of LMS, $t(118)=1.98$, $p=0.051$. However, participants that did not accept to repeat the task in the laboratory ($M=46.48$, $SD=4.63$) had higher scores in physical LMS than those that presented ($M=43.79$, $SD=7.02$).

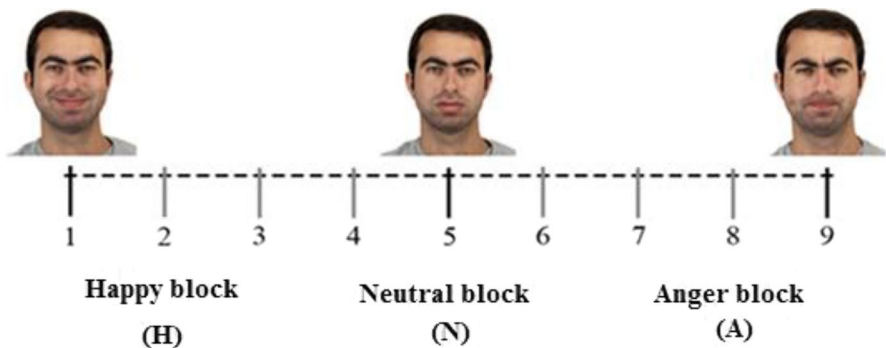


Fig. 1 Graphic representation of the created stimulus: a continuum of nine levels of transformed faces. Note. We grouped the image levels into three groups: Happy block, Levels 1 to 3; Ambiguous block, Levels 4 to 6; and, Angry block, Levels 7 to 9

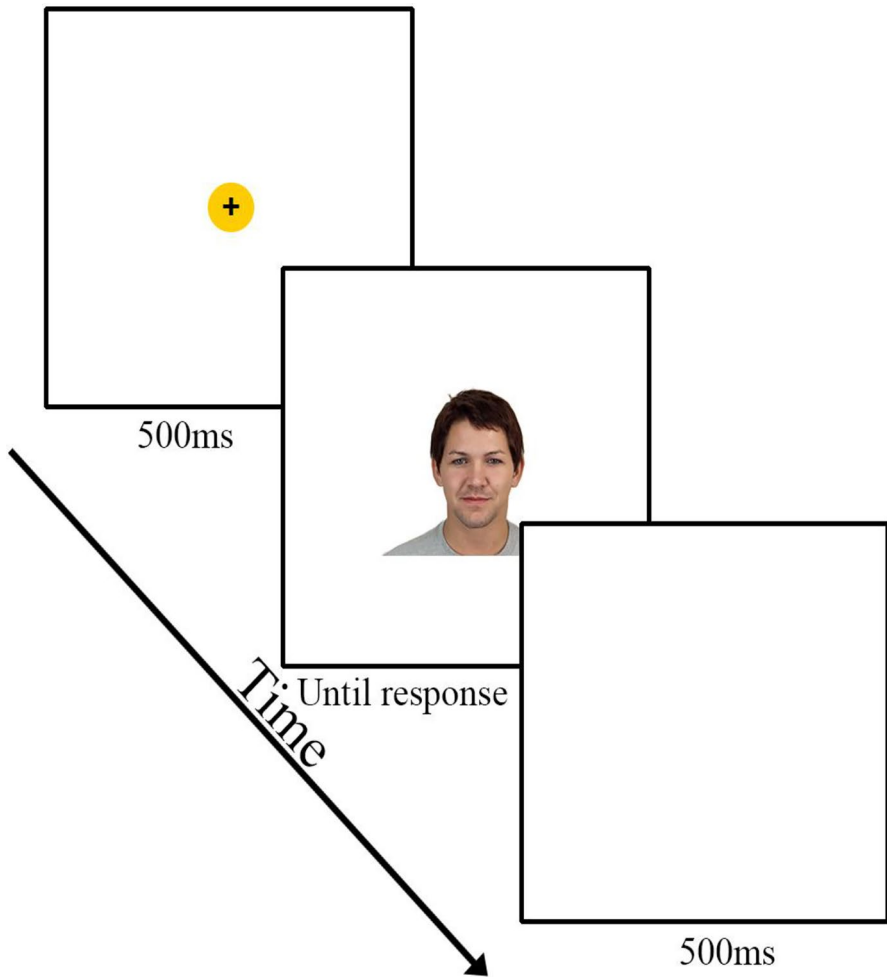


Fig. 2 Flow diagram of the interpretation bias task in the interpretation of the value of the faces. Note. First, to catch the participant's attention the "+" symbol is shown. Second, one of the 72 face images will appear until the participant responds by pressing the "N" or "P" keys on the keyboard: the "N" key for a face with a negative expression or the "P" for a face with a positive one

These differences were significant, $t(118)=2.13$, $p=0.035$. The study was approved by the Ethics Committee of the [Masked].

Indices and Data Analysis

The main index of the interpretation bias task was the frequency of all responses marked as negative ($F-$). Depending on the image type, the other three frequency indexes were calculated: $F(H-)$ was the number of happy faces (face levels 1–3) marked as negative, $F(N-)$ was the number of neutral faces (face levels 4–6) marked

as negative, and $F(A-)$ was the number of angry faces (face levels 7–9) marked as negative. The rest of the indices were based on RT, and when calculating them the data from the cases that were $\pm 3 SD$ from the mean were discarded. This cutoff was applied to each stimulus group (Fig. 1). $RT(H)$ was the mean of RT for happy faces (face levels 1–3). $RT(N)$ was the mean of RT for neutral faces (face levels 4–6). In the relation of this index, two sub-indices were created: $RT(N+)$, with only items marked as positive, and $RT(N-)$, with only negative ones. $RT(A)$, were the mean of RT for anger (face levels 7–9). Finally, $RT(Tot)$ consisted of the mean of all RT trials. Given the low number of negative responses for happy faces (levels 1 to 3) and positive responses for negative faces (levels 7 to 9), it was not possible to calculate the specific RT based on a negative or positive response for those instances.

The main analysis was carried out with SPSS® 24: Rho Spearman correlations (r_s) and Student t -test. To correct for multiple comparisons of the t -test, the Benjamini–Hochberg correction (false discovery rate, FDR) of p values was applied (Benjamini & Hochberg, 1995). To measure the effect sizes Cohen's d was calculated. To measure internal reliability Cronbach's alpha was used. Cronbach's alpha was calculated using data of the responses of each image (coded as "0" when positive, and coded as "1" when negative). For the comparison between groups (e.g., low and high SA), a group with one-third of the participants with lower scores and another group with one-third of those with higher scores was formed. Cutoffs were $x \leq 14$ for depression and $x \leq 35$ for SA in the low symptom group and $x \geq 25$ and $x \geq 47$ in the high symptom group, respectively. To detect outliers in general task performance, we established the criteria of $\pm 3 SD$ in one of the indices of the task (1 outlier).

Results

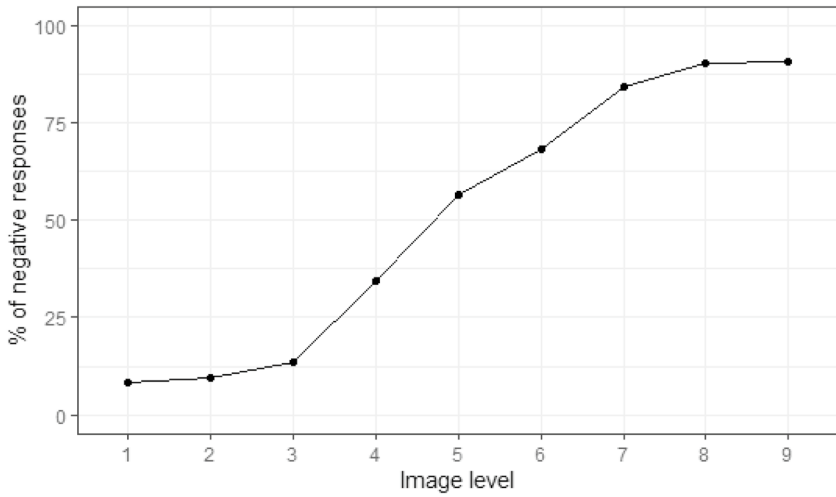
Interpretation Bias Task: Descriptive Analysis and Test–Retest Reliability

As anticipated, only 10.38% of the positive images were interpreted as negative, whereas only 11.46% of the negative faces were interpreted as positive (Fig. 3A). Importantly, 53.21% of the block of ambiguous faces were evaluated negatively. The participants needed less time to interpret happy faces as positive ($M=984.61$, $SD=345.73$) than to interpret the neutral block as positive ($M=1297.23$, $SD=522.17$) and the difference was significant, $t(854)=23.73$, $p<0.001$, $d=0.71$ (Fig. 3B). Similarly, negative faces were allocated faster to the negative condition ($M=1042.85$, $SD=427.94$) than neutral faces to the negative condition ($M=1237.28$, $SD=563.59$). This difference was also significant, $t(853)=18.05$, $p<0.001$, $d=0.39$.

In terms of reliability, the Cronbach alpha was 0.91. Regarding test–retest reliability, the main index, the number of face images marked as negative ($F-$), had a moderate test–retest correlation, $r_s(82)=0.59$, $p<0.001$ (Fig. 4). The number of face images marked as negative in angry faces, $F(A-)$, had the lowest test–retest coefficient, $r_s(82)=0.29$, $p<0.001$, but the test–retest correlations of the number of images marked as negative with neutral faces and with happy faces were moderate,

A

Percentage of negative responses by image level.



B

Reaction time by image level and response type.

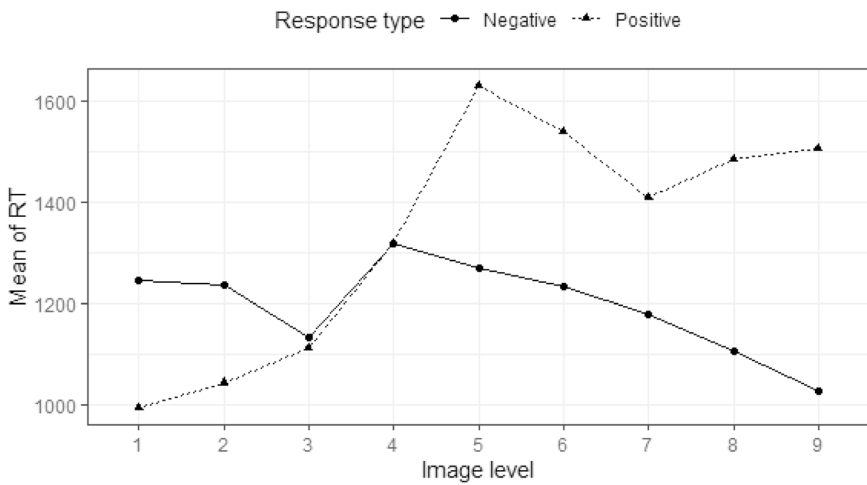


Fig. 3 The distribution of responses to the interpretation bias task. **A** Percentage of negative responses by image level. **B** Reaction time by image level and response type

respectively 0.57, $p < 0.001$ and 0.41, $p < 0.001$. The test–retest correlation for the indices based on RT was also moderate and between 0.55 and 0.69. The rest of the test–retest correlations followed a similar trend (Table 1). Regarding the correlations measured at the same time, high correlations were found between the indices based on RT in both waves. Specifically, in wave 1, the correlation coefficients between indices ranged between 0.78 and 0.96.

Association of Interpretation Bias with Depression, SA, and LMS

Table 2 shows the correlation matrix between the task indices and the rest of the variables. A positive and significant relationship was observed between the main index, $F-$, and depression and LMS. This trend was also observed for the number of neutral images interpreted as negative, that is, $F(N-)$ also showed a positive and significant correlation with depression, SA, and LMS. $F(A-)$ followed the same trend, except for depression, where the correlation was not significant. In the case of depressive symptoms, more symptoms were significantly associated with more negative interpretation of happy faces. Regarding RT based indexes, symptoms of depression were associated with a higher RT interpreting happy faces and social LMS was associated with $RT(N+)$ and $RT(H)$. However, the correlation coefficients in all cases were very low ($r \leq 0.22$).

Mean Differences in Interpretation Bias in SA and Depression

When creating subgroups (upper vs. lower tercile) for SA, there were statistically significant differences in $F(N-)$, $t(577) = 3.07$, $p = 0.002$, and $d = 0.26$; $F(A-)$,

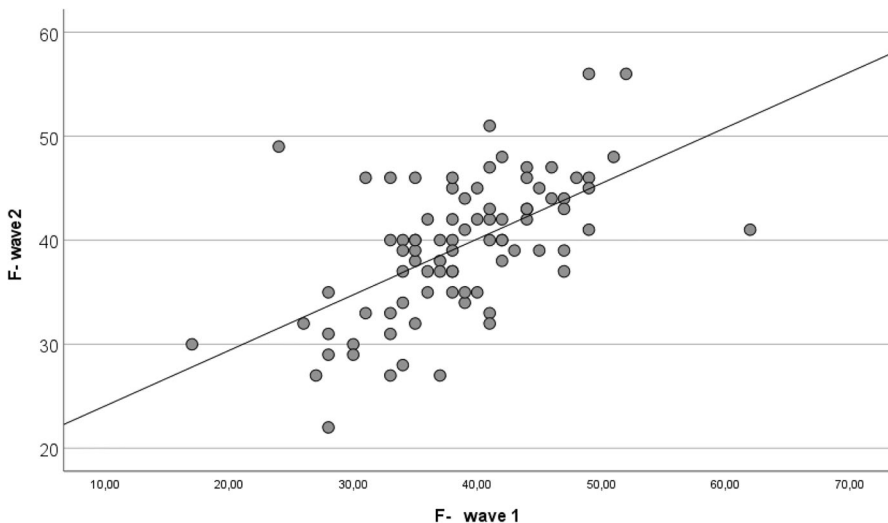


Fig. 4 Dispersion diagram between the numbers of faces interpreted as negative at wave 1 and wave 2. Note. $r_s(82) = .59$, $p < .001$, $r^2 = .34$. The sample that answered in wave 2 is shown. The elimination of the outliers was done in conjunction with wave 2. Therefore, the cases that might appear to be outliers are not in relation to the total sample

Table 1 Cross-sectional and test–retest Rho Spearman correlations between the task indices

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1. <i>F(H-)</i> T1																				
2. <i>F(N-)</i> T1	.40***																			
3. <i>F(A-)</i> T1	-.06	.46**																		
4. <i>F-T1</i> T1	.51***	.94***	.57**																	
5. <i>RT(H)</i> T1	.07*	-.05	-.04	-.03																
6. <i>RT(N+)</i> T1	-.07	-.10**	.06	-.07*	.78**															
7. <i>RT(N)</i> T1	-.14***	-.26***	-.06	-.25**	.81**	.90**														
8. <i>RT(N-)</i> T1	-.19***	-.33***	-.15**	-.33**	.74**	.75**	.93**													
9. <i>RT(A)</i> T1	-.14***	-.36***	-.24**	-.36**	.78**	.75**	.87**	.86**												
10. <i>RT T1</i> T1	-.08*	-.24***	-.12**	-.23**	.90**	.87**	.96**	.91**	.94**											
11. <i>F(H-)</i> T2	.41***	.20	.09	.26*	-.09	-.22*	-.25*	-.26*	-.13	-.18										
12. <i>F(N-)</i> T2	.07	.57***	.38***	.56***	-.04	-.02	-.13	-.16	-.25*	-.16	.36**									
13. <i>F(A-)</i> T2	.02	.38***	.29**	.37**	.06	.05	-.04	-.07	-.11	-.03	.15	.48***								
14. <i>F-T2</i> T2	.15	.58***	.39***	.59***	-.04	-.05	-.16	-.19	-.25*	-.17	.51**	.96***	.60***							
15. <i>RT(H)</i> T2	.03	-.11	-.15	-.13	.63***	.58***	.60***	.58***	.59***	.63**	-.03	-.17	.04	-.16						
16. <i>RT(N+)</i> T2	-.07	.00	-.02	-.03	.67***	.65***	.67***	.64***	.57***	.67***	-.24*	-.09	.16	-.09	.80***					
17. <i>RT(N)</i> T2	-.17	-.14	-.11	-.18	.62***	.64***	.69***	.67***	.61***	.68**	-.36**	-.32**	-.04	-.33**	.81***	.89***				
18. <i>RT(N-)</i> T2	-.14	-.20	-.13	-.23*	.62***	.63***	.69***	.69***	.63***	.68**	-.33**	-.32**	-.08	-.34**	.80***	.78***	.96***			

Table 1 (continued)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
19. RT(A) T2	-.10	-.30**	-.27*	-.34**	.57***	.55***	.62***	.60***	.65***	.65***	-.24*	-.46***	-.14	-.44***	.84***	.71***	.84***	.83***		
20. RT T2	-.12	-.20	-.19	-.24*	.64***	.64***	.69***	.67***	.65***	.70***	-.24*	-.33**	-.06	-.34**	.91***	.85***	.96***	.94***	.93***	
<i>n</i>	864	864	864	864	855	864	855	864	864	864	84	84	84	84	84	84	84	84	84	84
<i>M</i>	2.49	12.77	21.25	36.51	988.32	1297.23	1235.21	1238.51	1044.62	1088.05	1.49	15.04	22.87	39.39	921.60	1229.59	1134.37	1124.81	950.92	1002.12
<i>SD</i>	3.64	5.35	3.83	9.27	354.29	522.17	485.08	564.40	413.80	384.86	1.98	4.74	1.74	6.59	225.91	416.44	385.32	460.25	285.74	277.62

T1 = baseline; T2 = retest; RT(H) = the mean of RTs for the images of the happy block; $F(H-)$ = the frequency of happy faces classified as negative; $F(N-)$ = the frequency of neutral faces classified as negative; $F(A-)$ = the frequency of anger faces classified as negative; $RT(N+)$ = the mean of RTs for the images of the neutral block that were classified as positive; $RT(N)$ = the mean of RTs for the images of the neutral block; $RT(N-)$ = the mean of RTs for the images of the neutral block that were classified as negative; $RT(A)$ = the mean of RTs for the images of the neutral block; $RT(Tot)$ = the overall the mean of RTs of the task

* $p < .05$. ** $p < .01$. *** $p < .001$

Table 2 Rho Spearman correlation matrix between task indices, SAS-A, CES-D, and LMSQ-R

	F(H-)	F(N-)	F(A-)	F-	RT(H)	RT(N+)	RT(N)	RT(N-)	RT(A)	RT
CES-D	.12**	.15***	.03	.15***	.08*	.05	.02	-.00	.01	.03
SA-FNE	.00	.11**	.15**	.12**	.06	.07*	.04	.00	.02	.04
SA-New	.02	.10**	.14**	.11**	.01	.03	.00	-.03	-.02	-.00
SA-Gen	.08*	.10**	.07*	.10**	.06	.05	.04	.02	.02	.04
SA-Tot	.03	.11**	.14**	.12**	.05	.05	.02	-.01	.00	.03
LMS-Phy	.01	.16***	.12***	.15***	.04	.04	.01	-.04	-.02	.01
LMS-Soc	.04	.18***	.22***	.19***	.07*	.09**	.03	-.01	-.03	.02
LMS-Tot	.03	.19**	.20***	.19***	.06	.08*	.02	-.03	-.03	.02

SA-FNE=the Fear of Negative Evaluation subscale of the SAS-A scale; SAS-New=Social Avoidance and Distress in New Situations subscale of the SAS-A scale; SAS-Gen=Social Avoidance and Distress in General Situations subscale of the SAS-A scale; LMS-Soc=social dimension of The Looming Cognitive Style; LMS-phy=psychical dimension of the Looming Cognitive Style; LMC-Tot=global score of the Looming Cognitive Style

* $p < .05$. ** $p < .01$. *** $p < .001$

$t(536) = 3.82$, $p < 0.001$, and $d = 0.32$; and $F-$, $t(561) = 2.74$, $p = 0.006$, and $d = 0.23$, in which the group high in SA marked significantly more images as negative (Table 3A). For depression, we found that the high depression group scored higher on $F(H-)$, $t(544) = 2.84$, $p = 0.005$, and $d = 0.24$; $F(N-)$, $t(565) = 4.41$, $p < 0.001$, and $d = 0.37$; $F-$, $t(565) = 3.87$, $p < 0.001$, and $d = 0.32$; $RT(H)$, $t(565) = 2.17$, $p = 0.03$, and $d = 0.18$; and $RT(N+)$, $t(523) = 2.60$, $p = 0.01$, and $d = 0.22$ than the low depression group (Table 3B). That is, the high depression group marked more happy and neutral images as negative. They also needed more time to interpret happy faces and to mark neutral faces as positive.

Discussion

The objectives of this study were, first, to determine test–retest reliability and sources of validity, and second, to analyze the relationship between the indexes—those based on the negative interpretation of faces and those based on time response—of the interpretation bias task using ambiguous faces and symptoms of SA, depression, and LMS. The main findings were as follows: (1) the test–retest reliability of the main index of the task (i.e., the number of faces interpreted as negative) was moderate and the data supported the validity of the task through its association with SA, depression, and LMS; (2) SA, depressive symptoms, and LMS were associated with the number of ambiguous faces marked as negative.

The Task: Performance, Test–Retest Reliability, and Validity

The first objective explored the basic functioning, the reliability (specially, test–retest) of the task, and its validity. Regarding the basic functioning of the task, it was as expected. Thus, as the images approached anger, the participants marked

Table 3 Mean comparison in task indices between subgroups according to symptoms

A. Subgroups according to social anxiety symptoms								
	Low (<i>n</i> =292)		High (<i>n</i> =287)					
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>t</i>	<i>p</i>	<i>BH p</i>	<i>d</i>
<i>F</i> (<i>H</i> −)	2.88	4.44	2.43	3.26	−1.37	.170	.340	0.12
<i>F</i> (<i>N</i> −)	12.37	5.56	13.73	5.06	3.07**	.002	.010	0.26
<i>F</i> (<i>A</i> −)	20.67	4.40	21.90	3.26	3.82***	<.001	.010	0.32
<i>F</i> −	35.92	10.24	38.06	8.51	2.74**	.006	.020	0.23
RT(<i>H</i>)	987.38	409.66	1013.88	346.82	0.84	.402	.636	0.07
RT(<i>N</i> +))	1267.16	510.56	1338.43	511.34	1.67	.096	.240	0.14
RT(<i>N</i>)	1218.78	499.71	1244.74	443.79	0.66	.509	.636	0.05
RT(<i>N</i> −)	1236.10	586.60	1219.96	506.18	0.35	.725	.725	0.03
RT(<i>A</i>)	1032.23	418.00	1050.44	415.50	0.52	.599	.666	0.04
RT	1078.35	407.79	1101.74	371.05	0.72	.471	.636	0.06
B. Subgroups according to depressive symptoms								
	Low (<i>n</i> =284)		High (<i>n</i> =283)					
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>t</i>	<i>p</i>	<i>BH p</i>	<i>d</i>
<i>F</i> (<i>H</i> −)	2.00	3.08	2.82	3.73	2.84**	.005	.018	0.24
<i>F</i> (<i>N</i> −)	11.82	5.14	13.75	5.24	4.41***	<.001	.006	0.37
<i>F</i> (<i>A</i> −)	21.17	4.15	21.38	3.73	0.62	.535	.589	0.05
<i>F</i> −	35.00	8.88	37.94	9.24	3.87***	<.001	.006	0.32
RT(<i>H</i>)	948.75	349.42	1011.53	338.82	2.17*	.030	.055	0.18
RT(<i>N</i> +))	1231.87	430.82	1341.52	560.47	2.6*	.010	.022	0.22
RT(<i>N</i>)	1200.24	433.26	1257.55	493.73	1.47	.142	.195	0.12
RT(<i>N</i> −)	1231.87	430.82	1341.52	560.47	2.6*	.010	.022	0.22
RT(<i>N</i> −)	1229.66	561.03	1246.96	555.84	0.37	.714	.714	0.03
RT(<i>A</i>)	1014.27	366.63	1055.63	428.70	1.24	.217	.265	0.10
RT	1053.06	348.00	1106.79	387.27	1.74	.083	.13	0.15

We used a cutoff of one-third of the participants with the lowest and highest scores. *F*(*H*−)=the frequency of happy faces classified as negative. *F*(*N*−)=the frequency of neutral faces classified as negative. *F*(*A*−)=the frequency of anger faces classified as negative. *F*− =the frequency of the images classified as negative. RT(*H*)=the mean of RTs for the images of the happy block; RT(*N*+) =the mean of RTs for the images of the neutral block that were classified as positive; RT(*N*)=the mean of RTs for the images of the neutral block; RT(*N*−)=the mean of RTs for the images of the neutral block that were classified as negative; RT(*A*)=the mean of RTs for the images of the neutral block; RT(tot)=the overall mean of RTs of the task. *BH p*=*p* value corrected for multiple comparison with Benjamini–Hochberg procedure

p* < .05. *p* < .01. ****p* < .001

a higher number of responses as negative. This trend is congruent with the findings of other studies (Jusyte & Schöenberg, 2014; Richards et al., 2002; Schofield et al., 2007). Also, the participants required less time when their answers were congruent with the image. That is, compared with the time needed to interpret neutral faces, the average RT was shorter when happy faces were marked as positive and angry

faces as negative. In other words, there seems to be a higher cost when interpreting ambiguous faces.

With regard to reliability, the overall internal reliability was optimal; therefore, it demonstrates the homogeneity between items (the negative responses to the images). Concerning the test–retest reliability of the task, the overall frequency index, $F-$, which measures the tendency to evaluate faces as negative, obtained a medium effect size test–retest correlation. As discussed in the introduction, none of the studies that used similar tasks had previously evaluated the stability of this type of measure (e.g., Garner et al., 2009; Maoz et al., 2016). However, compared to the dot probe, which is conceptually and methodologically the most similar task, the present study found higher stability in its main index than studies that used the traditional dot-probe methodology (Chapman et al., 2019) and remained at similar levels to those in a recent study that improved the stability indexes of the dot probe (Aday & Carlson, 2019). Nonetheless, because the dot probe is a different task, this comparison has to be interpreted with caution and can be unwarranted. Thus, this study provides a first indicator of test–retest reliability on an interpretation bias task based on ambiguous faces. The rest of the indexes based on the frequency of the images marked as negative showed low to moderate reliability. Although the indexes that assess the frequency with which happy and angry faces were interpreted as negative resulted in the lowest test–retest correlation, the index based on a negative interpretation of neutral faces showed a moderate test–retest correlation and a very high correlation with the overall frequency index ($F-$). These data suggest that the reliability of the task is based mainly on the frequency of the number of responses marked as negative in neutral faces.

The analysis of the test–retest reliability of the RT rates revealed coefficients between moderate and high. In addition, the analysis of the correlation coefficients for different RT rates from wave 1 to wave 2 was very similar. For example, correlations from wave 1 to wave 2 of the same RT index (e.g., $RT(H)$ in wave 1 and $RT(H)$ in wave 2) showed no higher scores than the correlations of different rates in both waves (e.g., $RT(H)$ in wave 1 and $RT(N+)$ in wave 2). Due to these results, it might be assumed that these rates could be representations of the very same construct, such as “response time,” and that they might not differ from one another. Therefore, the methodology that was used may not be relevant for explaining individual differences in response time.

Regarding data acquisition, it is important to note that the task was performed online for the first wave, while the second time it was performed in a controlled manner in the laboratory. These data allow concluding that the reliability is good even when performing the task in an online format. In fact, the test–retest reliability would be expected to be higher if it had been evaluated in the two waves in a controlled manner in a laboratory. Future studies could compare reliability by taking into account the method of application (online vs. laboratory). For example, participants who used larger screens may have been able to capture more detail of the images compared to participants with smaller screens. Second, although it was not an objective of the study, it is important to remark that the test–retest coefficient is also a probe of the stability of the construct. Thus, regardless that interpretation

biases can be affected when experiencing anxiety, we can conclude that this construct has, at least, some grade of stability across time.

The present work also had a general aim to develop a task that captures interpretation biases that are relevant to symptoms of depression and social anxiety. The main task indexes—those based on the number of negative responses given to images—have been consistently associated with SA, depression and LMS. This consistency of relationship is a relevant source of the validity of the task. However, the effect sizes were low ($r \leq 0.20$; $d \leq 0.37$). This could be due to several reasons. First, the interpretation of ambiguous faces could be a characteristic that explains only a small part of this type of symptom. Second, different methods were used to assess interpretation biases (computerized tasks) and the other variables (self-reported instruments). Showing that correlations between related variables measured with different methods are low is common in the assessment of some psychological constructs (Morea & Calvete, 2020; Reinholdt-Dunne et al., 2013). Finally, for various reasons, such as the lack of ecological validity, computerized tasks may not fully capture the construct they are intended to measure. That is, the faces are not presented in a natural and ecological context. Future studies could continue this line of research with the goal to understand the best way to capture this type of bias, for example, analyzing if surprise faces (Mueller et al., 2020) capture better this bias. In addition, future studies could analyze which methodology best captures biases, for example, comparing a variety of methods in experimental research.

Relationship Across Interpretation Bias Task, SA, Depression, and LMS

Regarding the second objective, the results showed a positive association between the number of images marked as negative and symptoms of SA and depression. In addition, the group with high symptoms of depression and SA selected more images compared with the low symptoms group. The findings are in line with other studies that found such a bias toward negative interpretations of ambiguous faces in SA (Coles et al., 2008; Garner et al., 2009; Maoz et al., 2016; Schofield et al., 2007) and depression (Beevers et al., 2009; Joormann & Gotlib, 2006), but the effect sizes were low. This suggests that the interpretation bias of ambiguous faces explains a small part of SA and depression symptoms (~2–4%). More specifically in SA, other studies have shown that people with SA interpret ambiguous faces more negatively, expressing that faces indicate more rejection (Schofield et al., 2007), interpreting faces as less trustworthy (Gutiérrez-García & Calvo, 2016), or classifying more faces as angry (Joormann & Gotlib, 2006; Maoz et al., 2016). This trend was also observed in depression in other studies. In general, it has been found that people with depressive symptoms interpreted faces more negatively (Beevers et al., 2009; Joormann & Gotlib, 2006; Surguladze et al., 2004).

With regard to the association between the temporal cost (RT) to perform the interpretation and psychological symptoms, the data indicate that depressive symptoms are associated with a higher cost in interpreting ambiguous faces as positive and happy faces as positive or negative. These results are congruent with those of

another study in which a depressed group needed more time to interpret happy faces compared with the control group (Leppänen et al., 2004). This makes sense considering that these people have feelings of sadness, slower reasoning, and are biased to negative interpretations in different contexts (Everaert et al., 2017). Although a similar pattern was observed with the symptoms of SA—the more symptoms, the greater the temporal cost—the absence of statistical significance did not permit the inference of a relationship between SA and a higher cost in the interpretation of ambiguous faces as positive and the higher cost needed to interpret positive faces.

Another goal of the study was to examine the relationship between interpretation biases and LMS. The findings indicate that those who interpreted more ambiguous images as negative had a higher social LMS score. These results are in line with the study of Riskind and colleagues (2000) that associated the LMS with three different tasks of interpretation and memory bias. Contrary to what had been hypothesized, this relationship is present in both social and physical (e.g., heart attack sensations) scenarios. However, the time required to interpret neutral images as positive was only associated with the social LMS. These data support studies that relate LSM to more SA symptoms (Brown & Stopa, 2008; González-Díez et al., 2014, 2017). Conceptually, both LMS and interpretation bias are cognitive styles that drive negative interpretations of the context (Riskind et al., 2000). Although the data must be treated with caution due to the small size of the relationships between this vulnerability and the scores in the task, this is the first study that associated LMS with the interpretation bias assessed with ambiguous faces. The data from this study add more evidence to the relevance of the LMS in the relationship of variables associated with SA (Brown & Stopa, 2008; Calvete et al., 2016) and, more specifically, to cognitive biases (Riskind et al., 2000). This study opens the possibility for future studies to explore if LMS, as hypothesized, influences, or guides interpretation biases.

Limitation and Future Studies

Some limitations require discussion. The first problem was the high dropout rate for the laboratory task; of 130 invited students, only 84 completed the laboratory assessment (test–retest analysis). The sample that accepted the laboratory assessment could have certain characteristics; for example, people who agreed to participate were able to be more interested in the study and, therefore, follow the instructions more enthusiastically in both online and laboratory contexts. In the case of different characteristics between people that accepted participating and those who did not accept can be a problem in the generalization of the results. Second, the first wave was completed entirely online, so it is unknown whether or not the participants followed the guidelines indicated to carry out the task (e.g., not performing any other action that could be distracting). Likewise, one problem with online assessments, in general, is that it is hard to control the diversity of the devices that are used. For example, participants who used larger screens may have been able to capture more detail of the images compared to participants with smaller screens. Finally, although this was not an objective of the study, due to the variability between studies in the

task methodologies, it is not possible to conclude whether the changes made in the present study improved the capture of the interpretation bias. To do that, an experimental study comparing different methodologies is needed. Thus, considering the limitations, future studies could analyze the reliability in both online and laboratory contexts (both waves in online form or both in the laboratory). With the objective of answering other questions related to the use of tasks that measure interpretation biases with ambiguous faces, future studies also could compare different types of tasks with different stimuli or question types (e.g., emotion identification vs. face valence).

Conclusion

People with greater SA symptoms and greater depressive symptoms interpreted more morphed faces as negative. Moreover, the interpretation biases of ambiguous faces were associated with LMS. Finally, one of the main contributions is related to the examination of the test–retest reliability of the emotion-recognition task. The results provide evidence of the validity and reliability of the task. Thus, the task could be used over time, for example, for longitudinal studies or to assess the effects of an intervention on cognitive biases. Although more studies are needed to examine the reliability of this type of task, the reliability and validity data are satisfactory and support that the task can be used longitudinally.

Author Contribution A.P. collaborated with the design of the study, developed the Interpretation of Ambiguous Faces Task, analyzed the data, wrote the paper, and prepared the tables and figures; S.C.M. collaborated with the design of the study, collaborated with the interpretation of the data, and collaborated with the writing of the study; E.C. designed the study, collaborated with the interpretation of the data, and collaborated with the writing of the study. All authors approved the final version of the manuscript for submission.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. This research was supported by a grant from Basque Government (PRE-2019-1-0034) to the first author and a grant from the Basque Government to the research team (IT1532-22).

Data Availability The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aday, J. S., & Carlson, J. M. (2019). Extended testing with the dot-probe task increases test–retest reliability and validity. *Cognitive Processing*, 20(1), 65–72. <https://doi.org/10.1007/s10339-018-0886-1>
- Adler, A. D., & Strunk, D. R. (2010). Looming maladaptive style as a moderator of risk factors for anxiety. *Cognitive Therapy and Research*, 34(1), 59–68. <https://doi.org/10.1007/s10608-008-9221-y>
- Amir, N., Foa, E. B., & Coles, M. E. (1998). Negative interpretation bias in social phobia. *Behaviour Research and Therapy*, 36(10), 945–957. <http://www.ncbi.nlm.nih.gov/pubmed/9714945>
- Bantin, T., Stevens, S., Gerlach, A. L., & Hermann, C. (2016). What does the facial dot-probe task tell us about attentional processes in social anxiety? A systematic review. *Journal of Behavior Therapy and Experimental Psychiatry*, 50, 40–51. <https://doi.org/10.1016/j.jbtep.2015.04.009>
- Beck, A. T. (1976). *Cognitive therapy and the emotional disorders*. Internaciona Universitat Press.
- Beck, A. T., & Dozois, D. J. A. (2011). Cognitive therapy: Current status and future directions. *Annual Review of Medicine*, 62(1), 397–409. <https://doi.org/10.1146/annurev-med-052209-100032>
- Beck, A. T., & Haigh, E. A. P. (2014). Advances in cognitive theory and therapy: The generic cognitive model. *Annual Review of Clinical Psychology*, 10(1), 1–24. <https://doi.org/10.1146/annurev-clinpsy-032813-153734>
- Beevers, C. G., Wells, T. T., Ellis, A. J., & Fischer, K. (2009). Identification of emotionally ambiguous interpersonal stimuli among dysphoric and nondysphoric individuals. *Cognitive Therapy and Research*, 33(3), 283–290. <https://doi.org/10.1007/s10608-008-9198-6>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, 57(1), 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Brown, M. A., & Stopa, L. (2008). The looming maladaptive style in social anxiety. *Behavior Therapy*, 39(1), 57–64. <https://doi.org/10.1016/j.beth.2007.04.004>
- Calvete, E., & Cardenoso, O. (1999). Creencias y síntomas depresivos: Resultados preliminares en el desarrollo de una Escala de Creencias Irracionales abreviada. [Beliefs and depressive symptoms: preliminary results in the development of an abbreviated Irrational Belief Scale.] *Anales De Psicología*, 15, 179–190. https://www.um.es/analesps/v15/v15_2pdf/05v98_15calvete.PDF
- Calvete, E., Riskind, J. H., Orue, I., & Gonzalez-Diez, Z. (2016). Recursive associations among maladaptive cognitions and symptoms of social anxiety and depression: Implications for sex differences. *Journal of Social and Clinical Psychology*, 35(10), 807–821. <https://doi.org/10.1521/jscp.2016.35.10.807>
- Chapman, A., Devue, C., & Grimshaw, G. M. (2019). Fleeting reliability in the dot-probe task. *Psychological Research Psychologische Forschung*, 83(2), 308–320. <https://doi.org/10.1007/s00426-017-0947-6>
- Chen, J., Short, M., & Kemps, E. (2020). Interpretation bias in social anxiety: A systematic review and meta-analysis. *Journal of Affective Disorders*, 276, 1119–1130. <https://doi.org/10.1016/j.jad.2020.07.121>
- Clark, D. M., & McManus, F. (2002). Information processing in social phobia. *Biological Psychiatry*, 51(1), 92–100. [https://doi.org/10.1016/S0006-3223\(01\)01296-3](https://doi.org/10.1016/S0006-3223(01)01296-3)
- Clark, D. M., & Wells, A. (1995). A cognitive model of social phobia. In *Social Phobia: Diagnosis, Assessment, and Treatment*, 41(68), 00022–00023.
- Clark, D. A., Beck, A. T., Alford, B. A., Bieling, P. J., & Segal, Z. V. (2000). Scientific foundations of cognitive theory and therapy of depression. *Journal of Cognitive Psychotherapy*, 14(1), 100–106. <https://doi.org/10.1891/0889-8391.14.1.100>
- 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. <https://doi.org/10.1080/02699930701685919>
- Dineen, K. A., & Hadwin, J. A. (2004). Anxious and depressive symptoms and children’s judgements of their own and others’ interpretation of ambiguous social scenarios. *Journal of Anxiety Disorders*, 18(4), 499–513. [https://doi.org/10.1016/S0887-6185\(03\)00030-6](https://doi.org/10.1016/S0887-6185(03)00030-6)
- Ekman, P., & Friesen, W. V. (2003). *Unmasking the face: A guide to recognizing emotions from facial clues*. Ishk.
- Everaert, J., Podina, I. R., & Koster, E. H. W. (2017). A comprehensive meta-analysis of interpretation biases in depression. *Clinical Psychology Review*, 58(September), 33–48. <https://doi.org/10.1016/j.cpr.2017.09.005>

- Everaert, J., Bronstein, M. V., Castro, A. A., Cannon, T. D., & Joormann, J. (2020). When negative interpretations persist, positive emotions don't! Inflexible negative interpretations encourage depression and social anxiety by dampening positive emotions. *Behaviour Research and Therapy*, *124*, 103510. <https://doi.org/10.1016/j.brat.2019.103510>
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>
- Garaizar, P., & Reips, U. D. (2019). Best practices: Two Web-browser-based methods for stimulus presentation in behavioral experiments with high-resolution timing requirements. *Behavior Research Methods*, *51*(3), 1441–1453. <https://doi.org/10.3758/s13428-018-1126-4>
- Garner, M., Baldwin, D. S., Bradley, B. P., & Mogg, K. (2009). Impaired identification of fearful faces in generalised social phobia. *Journal of Affective Disorders*, *115*(3), 460–465. <https://doi.org/10.1016/j.jad.2008.10.020>
- González-Díez, Z., Orue, I., Calvete, E., & Riskind, J. H. (2014). Propiedades psicométricas del Cuestionario de Estilo Disfuncional de Looming (LMSQ-R) en jóvenes españoles. *Psicothema*, *26*(2), 260–266. <https://doi.org/10.7334/psicothema2013.206>
- González-Díez, Z., Orue, I., & Calvete, E. (2017). The role of emotional maltreatment and looming cognitive style in the development of social anxiety symptoms in late adolescents. *Anxiety, Stress, & Coping*, *30*(1), 26–38. <https://doi.org/10.1080/10615806.2016.1188920>
- Gutiérrez-García, A., & Calvo, M. G. (2016). Social anxiety and perception of (un)trustworthiness in smiling faces. *Psychiatry Research*, *244*, 28–36. <https://doi.org/10.1016/j.psychres.2016.07.004>
- Gutiérrez-García, A., Fernández-Martín, A., Del Líbano, M., & Calvo, M. G. (2019). Selective gaze direction and interpretation of facial expressions in social anxiety. *Personality and Individual Differences*, *147*(April), 297–305. <https://doi.org/10.1016/j.paid.2019.04.034>
- Heimberg, R. G., Brozovich, F. A., & Rapee, R. M. (2014). A cognitive-behavioral model of social anxiety disorder. In *Social Anxiety* (pp. 705–728). Elsevier. <https://doi.org/10.1016/B978-0-12-394427-6.00024-8>
- Heuer, K., Lange, W. G., Isaac, L., Rinck, M., & Becker, E. S. (2010). Morphed emotional faces: Emotion detection and misinterpretation in social anxiety. *Journal of Behavior Therapy and Experimental Psychiatry*, *41*(4), 418–425. <https://doi.org/10.1016/j.jbtep.2010.04.005>
- Joormann, J., & Gotlib, I. H. (2006). Is this happiness I see? Biases in the identification of emotional facial expressions in depression and social phobia. *Journal of Abnormal Psychology*, *115*(4), 705–714. <https://doi.org/10.1037/0021-843X.115.4.705>
- Jusyte, A., & Schönenberg, M. (2014). Threat processing in generalized social phobia: An investigation of interpretation biases in ambiguous facial affect. *Psychiatry Research*, *217*(1–2), 100–106. <https://doi.org/10.1016/j.psychres.2013.12.031>
- Kappenman, E. S., Farrens, J. L., Luck, S. J., & Proudfit, G. H. (2014). Behavioral and ERP measures of attentional bias to threat in the dot-probe task: Poor reliability and lack of correlation with anxiety. *Frontiers in Psychology*, *5*, 1–9. <https://doi.org/10.3389/fpsyg.2014.01368>
- La Greca, A. M., & Lopez, N. (1998). Social anxiety among adolescents: Linkages with peer relations and friendships. *Journal of Abnormal Child Psychology*, *26*(2), 83–94. <https://doi.org/10.1023/A:1022684520514>
- Lee, J. S., Mathews, A., Shergill, S., & Yiend, J. (2016). Magnitude of negative interpretation bias depends on severity of depression. *Behaviour Research and Therapy*, *83*, 26–34. <https://doi.org/10.1016/j.brat.2016.05.007>
- Leppänen, J. M., Milders, M., Bell, J. S., Terriere, E., & Hietanen, J. K. (2004). Depression biases the recognition of emotionally neutral faces. *Psychiatry Research*, *128*(2), 123–133. <https://doi.org/10.1016/j.psychres.2004.05.020>
- Lira Yoon, K., & Zinbarg, R. E. (2007). Threat is in the eye of the beholder: Social anxiety and the interpretation of ambiguous facial expressions. *Behaviour Research and Therapy*, *45*(4), 839–847. <https://doi.org/10.1016/j.brat.2006.05.004>
- Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, *47*(4), 1122–1135. <https://doi.org/10.3758/s13428-014-0532-5>
- MacLeod, C., Grafton, B., & Notebaert, L. (2019). Anxiety-linked attentional bias: Is it reliable? *Annual Review of Clinical Psychology*, *15*(1), annurev-clinpsy-050718-095505. <https://doi.org/10.1146/annurev-clinpsy-050718-095505>

- Maoz, K., Eldar, S., Stoddard, J., Pine, D. S., Leibenluft, E., & Bar-Haim, Y. (2016). Angry-happy interpretations of ambiguous faces in social anxiety disorder. *Psychiatry Research*, *241*, 122–127. <https://doi.org/10.1016/j.psychres.2016.04.100>
- Morea, A., & Calvete, E. (2020). Validación del Cuestionario Breve de Funciones Ejecutivas Webexec: Pruebas de rendimiento y síntomas psicológicos [Validation of the Brief Questionnaire of Executive Functions Webexec: Performance Tests and Psychological Symptoms.]. *Revista de Psicología Clínica Con Niños y Adolescentes*, *7*(2). <https://doi.org/10.21134/rpcna.2020.07.2.7>
- Mueller, S. C., Unal, C., Saretta, M., Al Mughairbi, F., Gómez-Odrizola, J., Calvete, E., & Metin, B. (2020). Working memory and emotional interpretation bias in a sample of Syrian refugee adolescents. *European Child and Adolescent Psychiatry*, *0123456789*. <https://doi.org/10.1007/s00787-020-01656-8>
- Olivares, J., Ruiz, J., Hidalgo, M. D., García-López, L. J., Rosa, A. I., & Piqueras, J. A. (2005). Social Anxiety Scale for Adolescents (SAS-A): Psychometric properties in a Spanish-speaking population. *International Journal of Clinical and Health Psychology*, *5*(1), 85–97.
- Prieto-Fidalgo, Á., Miers, A. C., & Calvete, E. (2022). Interpretation biases in social scenarios and social anxiety: The role of safety behaviors. *Psicothema*, *23*(4), 489–497. <https://doi.org/10.7334/psicothema.2021.406>
- Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, *1*, 385. <https://doi.org/10.1063/1.4932809>
- Reinholdt-Dunne, M. L., Mogg, K., & Bradley, B. P. (2013). Attention control: Relationships between self-report and behavioural measures, and symptoms of anxiety and depression. *Cognition & Emotion*, *27*(3), 430–440. <https://doi.org/10.1080/02699931.2012.715081>
- Richards, A., French, C. C., Calder, A. J., Webb, B., Fox, R., & Young, A. W. (2002). Anxiety-related bias in the classification of emotionally ambiguous facial expressions. *Emotion*, *2*(3), 273–287. <https://doi.org/10.1037/1528-3542.2.3.273>
- Riskind, J. H., & Williams, N. L. (2005). The looming cognitive style and generalized anxiety disorder: Distinctive danger schemas and cognitive phenomenology. *Cognitive Therapy and Research*, *29*(1), 7–27. <https://doi.org/10.1007/s10608-005-1645-z>
- Riskind, J. H., Williams, N. L., Gessner, T. L., Chrosniak, L. D., & Cortina, J. M. (2000). The looming maladaptive style: Anxiety, danger, and schematic processing. *Journal of Personality and Social Psychology*, *79*(5), 837–852. <https://doi.org/10.1037/0022-3514.79.5.837>
- Ruiz-Grosso, P., Loret de Mola, C., Vega-Dienstmaier, J. M., Arevalo, J. M., Chavez, K., Vilela, A., Lazo, M., & Huapaya, J. (2012). Validation of the Spanish Center for Epidemiological Studies Depression and Zung Self-Rating Depression Scales: A comparative validation study. *PLoS ONE*, *7*(10), e45413. <https://doi.org/10.1371/journal.pone.0045413>
- Sanchez, A., Everaert, J., De Putter, L. M. S., Mueller, S. C., & Koster, E. H. W. (2015). Life is ... great! Emotional attention during instructed and uninstructed ambiguity resolution in relation to depressive symptoms. *Biological Psychology*, *109*, 67–72. <https://doi.org/10.1016/j.biopsycho.2015.04.007>
- Schofield, C. A., Coles, M. E., & Gibb, B. E. (2007). Social anxiety and interpretation biases for facial displays of emotion: Emotion detection and ratings of social cost. *Behaviour Research and Therapy*, *45*(12), 2950–2963. <https://doi.org/10.1016/j.brat.2007.08.006>
- Schultz, L. T., & Heimberg, R. G. (2008). Attentional focus in social anxiety disorder: Potential for interactive processes. *Clinical Psychology Review*, *28*(7), 1206–1221. <https://doi.org/10.1016/j.cpr.2008.04.003>
- Shiroma, P. R., Thuras, P., Johns, B., & Lim, K. O. (2016). Facial recognition of happiness among older adults with active and remitted major depression. *Psychiatry Research*, *243*, 287–291. <https://doi.org/10.1016/j.psychres.2016.06.020>
- Soto, F. A., Stewart, R. A., Hosseini, S., Hays, J., & Beevers, C. G. (2021). A computational account of the mechanisms underlying face perception biases in depression. *Journal of Abnormal Psychology*, *130*(5), 443–454. <https://doi.org/10.1037/abn0000681>
- Surguladze, S. A., Young, A. W., Senior, C., Brébion, G., Travis, M. J., & Phillips, M. L. (2004). Recognition accuracy and response bias to happy and sad facial expressions in patients with major depression. *Neuropsychology*, *18*(2), 212–218. <https://doi.org/10.1037/0894-4105.18.2.212>