Social anxiety is characterized by biased learning about performance and the self

Leonie Koban[1,2], **Rebecca Schneider**[*,1], **Yoni K. Ashar**[*,1,2], **Jessica R. Andrews-Hanna**[1,2], **Lauren Landy**[1], **David A. Moscovitch**[3], **Tor D. Wager**[1,2], and **Joanna J. Arch**[2]

[1]Institute of Cognitive Science, University of Colorado Boulder, 345 Muenzinger, Boulder, CO 80309

[2]Department of Psychology and Neuroscience, University of Colorado Boulder, 345 Muenzinger, Boulder, CO 80309

[3]Department of Psychology and Centre for Mental Health Research, University of Waterloo, 200 University Avenue West, Waterloo, ON, N2L 3G1, Canada

Abstract

People learn about their self from social information, and recent work suggests that healthy adults show a positive bias for learning self-related information. In contrast, social anxiety disorder (SAD) is characterized by a negative view of the self, yet what causes and maintains this negative self-view is not well understood. Here we employ a novel experimental paradigm and computational model to test the hypothesis that *biased social learning* regarding self-evaluation and self-feelings represents a core feature that distinguishes adults with SAD from healthy controls. Twenty-one adults with SAD and 35 healthy controls (HC) performed a speech in front of three judges. They subsequently evaluated themselves and received performance feedback from the judges, and then rated how they felt about themselves and the judges. Affective updating (i.e., change in feelings about the self over time, in response to feedback from the judges) was modeled using an adapted Rescorla-Wagner learning model. HC demonstrated a positivity bias in affective updating, which was absent in SAD. Further, self-performance ratings revealed group differences in learning from positive feedback—a difference that endured at an average of 1 year follow up. These findings demonstrate the presence and long-term endurance of positively biased social learning about the self among healthy adults, a bias that is absent or reversed among socially anxious adults.

Keywords

Social Anxiety Disorder; Learning; Social Emotions; Positivity Bias; Feedback Monitoring

How we perceive and feel about ourselves is a central aspect of human experience and an important determinant of subjective well-being. Poor self-esteem or self-image is characteristic of a wide range of mental health disorders. Social anxiety disorder (SAD)—one of the most common psychological disorders (Kessler, Chiu, Demler, & Walters, 2005)—is characterized by excessive fear and avoidance of social situations (Kessler et al., 2005) and by a persistently negative view of the self (Hirsch, Clark, Mathews, & Williams, 2003; Moscovitch, 2009). SAD has detrimental consequences not only for the social life of affected individuals, but also for their academic success, employment, economic status, and overall quality of life (Safren, Heimberg, Brown, & Holle, 1996; Stein & Kean, 2000; Wittchen, Fuetsch, Sonntag, Müller, & Liebowitz, 2000).

A fundamental question is why SAD and associated negative feelings towards the self persist even in the face of regular exposure to social situations without overt negative consequences (Clark, 2005). The current study employs a novel social feedback task and computational modeling approach to test the hypothesis that negatively biased learning from social feedback plays a critical role in maintaining beliefs regarding social deficits and negative feelings towards the self within SAD. A recent study (Korn, Prehn, Park, Walter, & Heekeren, 2012) has demonstrated that self-evaluation can be influenced by social feedback - how we think that others perceive us. For example, if others tell us that we are intelligent, we will adjust our self-image in this direction and think of ourselves as more intelligent. Healthy adults show a *positivity bias* in how they interpret and learn from social feedback: they incorporate social feedback into their self-image to a greater extent when the feedback is positive than when it is negative (Korn et al., 2012; Sharot, Korn, & Dolan, 2011; Taylor & Brown, 1988). These findings are in line with more general positive or optimistic biases that may be absent or reduced in depressed individuals (Garrett et al., 2014; Taylor & Brown, 1988). However, it is unknown whether the absence of this positivity bias in a clinical population extends to social learning about the *self* and one's ability to perform in social contexts.

Relatedly, no research to date has addressed how positive and negative social feedback dynamically affect how people feel about themselves, that is, how state self-esteem (the affective component of the self, Brown, 1993) fluctuates as a function of integrating social performance feedback over time (Brown, 2010; Swann, Griffin, Predmore, & Gaines, 1987). The sociometer theory (Leary, Tambor, Terdal, & Downs, 1995) proposes that self-esteem depends on our beliefs about how others evaluate us and on our subjective sense of being accepted versus rejected by others. Although state self-esteem fluctuates in response to social evaluations from others (Eisenberger, Inagaki, Muscatell, Haltom, & Leary, 2011), it is not known whether healthy people are differentially sensitive to positive versus negative feedback, and how this might be altered in SAD.

Cognitive models emphasize negative self-perception as a fundamental problem in SAD that lies at the core of the disorder (Clark & Wells, 1995; Hofmann, 2007; Moscovitch, 2009;

Rapee & Heimberg, 1997). Indeed, SAD is characterized by excessive preoccupation with perceived self-flaws in social competence, signs of anxiety, and/or physical appearance that could become exposed for scrutiny and potential criticism by evaluative others (Moscovitch et al., 2013). One influential cognitive theory suggests that fear of social situations in SAD leads to excessive attention to (internal) bodily arousal and anxiety, and reduced processing of (external) social feedback (Clark, 2005). Consequently, for those with SAD, social situations are intrinsically aversive and misconceptions about social interactions and self-impressions are not corrected by social feedback (Clark, 2005).

The current study investigates how social learning influences feelings towards the self and one's social performance among both healthy and socially anxious adults. Specifically, we evaluate the extent to which these groups can be distinguished by distinct biases in *social learning*. Negatively biased social learning may represent a core online process in SAD, which could explain the persistence of a negative self-image despite repeated exposure to social situations (which can offer positive, corrective social feedback). Thus, we test two competing hypotheses: If SAD is characterized by reduced processing of external social information altogether, we would expect reduced learning in SAD compared to healthy controls (HC), regardless of feedback valence. Alternatively, if SAD is characterized by negatively biased social learning relative to healthy controls, we would expect SAD to reflect greater learning from negative relative to positive social feedback, and HC to reflect the opposite bias (see Figure 1A).

In the current study, 21 participants meeting DSM-IV diagnostic criteria for SAD and 35 healthy controls (HC) were asked to give a challenging speech (on a self-relevant topic) in front of three judges (Kirschbaum, Pirke, & Hellhammer, 1993) (Figure 1B). The judges rated their performance across 58 dimensions relevant to proposed core fears in SAD: signs of anxiety, competence, and appearance (Moscovitch, 2009). Following the speech, all participants completed a novel self-evaluation and social feedback task (Figure 1C). For each of the 58 feedback trials in this task, participants first rated their own performance on a particular dimension and then received feedback from the judges on that same dimension. For each trial, both sets of feedback (their own and the judges) remained on the screen at the same time, allowing participants to see any discrepancy between their own and the judges' performance ratings. They were then asked to rate how they felt about themselves and about the judges. In a second and third round of 58 trials completed 20 minutes later (T2) and, on average, 1 year later (T3), participants evaluated their performance again, this time without feedback from the judges.

Our novel task design allowed us to use a computational modeling approach to model dynamic updating of *feelings about the self* as a function of receiving positive and negative performance feedback from the judges. Thus, we predicted that healthy people would change how they feel about themselves more in response to positive social evaluative feedback, whereas adults with SAD would change how they feel about themselves more in response to negative feedback. We further tested how this *affective updating* predicted the subsequent occurrence of anxiety-related thoughts as assessed during spontaneous thought sampling, thus linking affective updating to verbal behaviors. Second, this task allowed us to quantify how self-ratings of one's speech performance were altered as a function of the

judges' feedback, both shortly after the speech and at follow-up, about 1 year later. Similarly, we predicted that adults with SAD would update evaluations of their speech performance to a greater extent following negative as compared to positive performance feedback whereas HC would update evaluations of their performance to a greater extent following positive relative to negative performance feedback.

Materials and Methods

Participants

Fifty-six participants were recruited from greater Boulder, Colorado with advertisements targeting socially anxious and healthy adults. All participants were first screened via email for initial inclusion and exclusion criteria, and subsequently clinically assessed using the MINI International Neuropsychiatric Interview for DSM-IV (Sheehan et al., 1998).

Screening—MINI phone interviews were administered by a trained doctoral student in clinical psychology or post-baccalaureate research assistant. Interviewers were extensively trained in the MINI, including achieving reliable diagnoses across 6 gold-standard training cases. The MINI is a brief, structured, validated clinical interview (Sheehan et al., 1998) that assesses anxiety, mood, substance, and eating disorders, and screens for suicidality and psychosis. In the current study, the MINI was enhanced with more detailed diagnostic criteria for the anxiety disorders, as piloted in previous studies (Roy-Byrne et al., 2010).

Inclusion and exclusion criteria—All participants had to be 18–40 years old, physically healthy, and not pregnant. Participants in the SAD group ($n = 21$) met DSM-IV criteria for social anxiety disorder on the MINI (including fear of public speaking), without meeting criteria for a current major depressive episode, current dysthymia, or any other Axis I disorders (except other anxiety disorders). Participants in the healthy control group (HC, $n = 35$) were required to not meet diagnostic criteria for any of the psychological disorders assessed by the MINI. For both groups, additional exclusion criteria included current or recent self-harm, current suicidal ideation, intent, or plan, suicide attempt in the past 5 years, current heavy substance use (defined as meeting criteria for substance abuse or dependence as assessed by the MINI, or daily use), recent head injuries or neurologic disorders. Any prescription medication use had to be stable. Participants were instructed to not take any other medication on the day of the study, prior to the laboratory session.

Table 1 presents participant characteristics. As there were more female participants in the SAD compared to the HC group, all between-group analyses were repeated with sex as a covariate to exclude the possibility that group effects were driven by sex differences. In addition, we performed female-only analyses (see Supplemental Materials), which demonstrated that the female-only sample showed the same group differences as the total sample.

Questionnaires—In the beginning of the experimental session, participants completed a battery of clinical and personality questionnaires in randomized order, including the Overall Anxiety Severity and Impairment Scale (OASIS, Norman, Hami Cissell, Means-Christensen, & Stein, 2006) and State Trait Anxiety Inventory (STAI, Spielberger, 1970) to measure

anxiety, the Patient Health Questionnaire (PHQ-9, Spitzer, Kroenke, Williams, Group, & others, 1999) to assess depression, the Fear of Negative Evaluation Scale (FNE, Watson & Friend, 1969), Social Interaction Anxiety Scale (SIAS, Mattick & Clarke, 1998), and Negative Self-Portrayal Scale (NSPS, Moscovitch & Huyder, 2011) to measure aspects of social anxiety, the Self-Compassion Scale (SCS, Neff, 2003) to measure self-compassion, Acceptance and Action Questionnaire (AAQ-II, Bond et al., 2011) to measure psychological inflexibility and experiential avoidance, the Rumination and Reflection Questionnaire (RRQ, Trapnell & Campbell, 1999) to measure rumination, and basic demographic information. See Table 1 for mean questionnaire scores in the two experimental groups.

Procedures

Speech task—The speech task was adapted from the Trier Social Stress Test paradigm (Kirschbaum et al., 1993). Participants were given 3 min to mentally prepare a speech (without writing anything down) on their perfect job and why they were well suited for it, to be performed in front of a panel of judges. Then, three confederates in the role of speech performance 'judges' with each slightly negative, neutral, or slightly positive ('warm') attitudes entered the room and the participant was asked to speak for 5 minutes. During the speech, judges used laptops to rate participants on 58 items that broadly emphasized the three core self-focused fears in SAD, namely competence, appearance, and signs of anxiety (Moscovitch, 2009) (see Table S1 in the *Supplemental Materials* for the complete list of evaluative items). They were instructed to not communicate any strong signs of approval, feedback, or strong facial expressions, and to rate as accurately as possible on a 0–100 scale (0 representing worst possible and 100 the best possible performance, subsequently transformed to values between 0–1) while maintaining their roles (i.e., the negative judge maintained a stoic face and rated more critically, the positive judge nodded occasionally and rated more positively, the neutral judge rated moderately and maintained a neutral expression). This ensured that each participant was assigned a broad range of ratings.

Self-evaluation and social feedback task (T1)—After the speech, an optimization algorithm selected one of the judges' ratings per evaluative item in a manner that matched the valence of the selected feedback to a uniform distribution (see Supplemental Materials). This allowed us to provide each participant with authentic feedback on all 58 items, and at the same time provide similar amounts of feedback across the positive-to-negative visual analog scale for each participant. As presented in Figure 1C, each trial of the feedback task started with the presentation of the evaluative item (e.g. "My speech was engaging"), after which the participant had to rate his/her own performance on the visual analog scale (VAS) with anchors 'applies not at all' to 'applies very much' (corresponding to values ranging from 0 to 1). Immediately following their own rating, participants were presented with the judges' feedback (for 3s) on the same item (e.g., "Her speech was engaging"), while their own rating remained on the screen. Participants did not know which judge provided which rating, but they were instructed prior to the task that ratings were randomly picked from any of the judges, which differed in their opinion on how to best give a speech. Then participants used a VAS to rate how they felt about themselves ("*How do you feel about yourself overall?*", anchored at 'very bad and 'very good'), and how they felt about the judges ("*How do you feel about the judges overall?*").

Self-evaluation follow-up (T2 and T3)—Approximately 20 min after completing the first round of ratings (T2), participants were asked to rate their performance again on the same 58 items, this time without receiving the judges' feedback and without subsequently rating their feelings about themselves and the judges. This allowed us to assess the extent to which judges' performance rating at T1 influenced participant's performance self-ratings at T2. To test for long-term changes in self-ratings as a function of T1 judges' feedback, participants were invited to rate their performance for a final time online (T3, via Qualtrics) at 7 to 17 months after the initial study session ($M = 11.7$ months, $SD = 2.8$ months, Range = 7.3 – 16.4 months).

Thought sampling—Participants were audio-recorded while describing what they currently were thinking on three occasions: 1) after being notified of the upcoming speech task, 2) after delivering the speech, and 3) after completing the self-evaluation and social feedback task (T1). At each occasion, participants were asked to speak aloud for one minute with no experimenters present in the room.

Analysis

Model fitting (affective updating)—To assess how emotions were dynamically altered as a function of social feedback during the task, we computationally modeled how social feedback (e.g., performance feedback from the judges) influenced subsequent *feelings about the self* using an adapted Rescorla-Wagner reinforcement learning model (Rescorla & Wagner, 1972). The main assumption of our model is that how people feel changes as a function of the value of the feedback they receive over time. Specifically, how people felt about themselves in a given trial t ($Feeling_{Self}(t)$) was modeled as a function of how they felt about themselves in the previous trial ($Feeling_{Self}(t-1)$) plus the difference term (affective prediction error, *APE*) between the previous feeling state and the valence of the feedback provided by the judges in the current trial ($V_{Feedback}(t)$), multiplied by a learning rate, a. For example, if a person feels very negatively about their self (e.g. previous rating of 0.2), but receives moderate feedback of 0.5, the APE would be positive (0.3) and the person would feel somewhat better than before as a consequence (between 0.2 and 0.5, depending on the learning rate a). We were specifically interested in differences regarding the integration of positive versus negative affective prediction errors, and thus we fitted separate learning rates for positive ($a_{SelfPos}$) and for negative ($a_{SelfNeg}$) *APEs*. These positive and negative individual learning rates constituted our variables of interest.

$$Feeling_{Self}(t) = Feeling_{Self}(t-1) + \begin{cases} \alpha_{SelfPos} APE & for & APE > 0 \\ \alpha_{SelfNeg} APE & for & APE < 0 \end{cases}$$
$$with \ APE = V_{Feedback}(t) - Feeling_{Self}(t-1)$$

Lower learning rates indicate a slower integration of social feedback into one's feeling about the self, whereas higher learning rates indicate faster integration of social feedback into one's feeling about the self, reflecting higher sensitivity or affective reactivity towards the judges' feedback (see *Supplemental Materials* Figure S2 for sample participants). For the SAD group, we predicted higher negative learning rates ($a_{SelfNeg}$) and lower positive

learning rates ($a_{SelfPos}$), that is, a tendency to more quickly incorporate negative judges' feedback into one's feelings about the self and to more slowly incorporate positive judges' feedback into one's feelings about the self, on a trial-by-trial, dynamic basis. For the HC group, we predicted the opposite. We employed the same updating model for participants' *feeling about the judges (Feeling_Judges)*, which allowed us to test whether group differences in learning rates were specific to feelings towards the self or generalized to feelings towards (evaluative) others.

Models were fitted to individual data using the Matlab Optimization toolbox, by minimizing the sum of square errors between modeled and data time course. All free parameters (initial value of *Feeling_Self*, *Feeling_Judges*, a_{pos}, a_{neg}) were constrained to values between 0 and 1. Individual model fit were compared to a simpler model with only one learning rate (for both positive and negative *APE*) using Bayesian model comparison (Bayesian Information Criterion, Schwarz, 1978). Time courses for the data and modeled time courses of two example participants are shown in *Supplemental Materials* Figure S2.

Adjustments in self-evaluation—The majority of initial participants (34 of 56, including 19 HC, 15 SAD) responded to the T3 survey. One SAD subject was excluded from the T3 analyses because their beta estimates were more than 3 SD below the mean, resulting in a final sample of 33 participants for this analysis. VAS ratings from the self-evaluation ratings ranged from 0 to 1 and were reverse-scored for negative statements. A multi-level GLM (Matlab code available at http://wagerlab.colorado.edu/tools) was used to obtain individual regression (beta) weights for the influence of feedback mismatch (*Eval*), defined as the difference between the *judges' feedback(T1)* and *self-evaluation(T1)*, on subsequent adjustments in self-evaluation, that is, on change in performance self-ratings from T1 to T2 and from T1 to T3. Feedback mismatch was positive when judges rated participants higher than participants rated themselves; feedback mismatch was negative when judges rated participants lower than participants rated themselves. Trials that reflected positive versus negative feedback mismatch were modeled separately, yielding separate regression (beta) weights for positive and negative individual learning from social feedback.

$$SelfEval(T2) = SelfEval(T1) + \beta_{pos}\Delta Eval \quad for \quad \Delta Eval > 0$$
$$SelfEval(T2) = SelfEval(T1) + \beta_{neg}\Delta Eval \quad for \quad \Delta Eval < 0$$

Importantly, distance to the mean (*SelfEval(T1)*-0.5) was added as a covariate to control for statistical regression to the mean.

Linguistic analysis of spontaneous thought content—Audio recordings of participants' spontaneous thought descriptions were transcribed, and transcriptions were analyzed for affective and anxiety-related semantic content using Linguistic Inquiry Word Count (LIWC2007) software (Pennebaker, Booth, & Francis, 2007). LIWC2007 references each word (independent of context) to an online dictionary of linguistic and psychological processes. We extracted the percentage of negative affect-related and anxiety-related words during the thought sampling phase following the feedback task (T1) for each participant, and used robust partial correlations to measure the relationship with biases in affective updating,

controlling for feedback content before the feedback task. We predicted a correlation between words related to negative affect and anxiety and negatively biased affective updating.

Results

Affective Updating

Model comparisons using the Bayesian Information Criterion (BIC, Schwarz, 1978) showed that individually fitted models that included separate learning rates for positive and negative *APE* in addition to the initial feeling state value outperformed simpler models that used only one learning rate per subject (i.e., combined positive and negative learning rates): for the *Feeling$_{Self}$* models, mean BIC values for the one- versus two-learning rate models were −217.7 and −712.9, respectively, $t(55) = 10.78$, $p < 0.001$, $d = 1.74$; for the *Feeling$_{Judges}$* models, BIC values were −184.1 and −710.1, $t(55) = 10.89$, $p < 0.001$, $d = 1.61$ (see *Supplemental Materials* Figure S3 for more information). This indicates that participants indeed learned differently from positive and negative feedback.

We next used a mixed-effects ANOVA to investigate how individual learning rates (as the dependent variables) differed between groups for positive and negative affective updating. In support of our hypothesis, the results showed a significant valence (positive *versus* negative *APE*) by affective rating type (feelings about self *versus* judges) by group (SAD *versus* HC) interaction, $F(1,54) = 16.25$, $p < 0.001$, partial $\eta^2 = 0.23$ (see Figure 2A), but no main effect of group on affective updating (across valence) and no main or interaction effects of sex (all Fs < 1, see *Supplemental Materials* Figure S4). Planned comparisons showed that both groups had higher positive than negative learning rates for feelings towards the judges, $F(1,54) = 6.51$, $p = 0.014$, partial $\eta^2 = 0.11$; that is, they updated their feelings about the judges more quickly and to a greater extent when the judges rated them more positively. In contrast, the SAD group updated their feelings about the *self* very differently than the HC group, as seen in a significant group by valence interaction, $F(1,54) = 13.55$, $p = 0.001$, partial $\eta^2 = 0.20$. As presented in Figure 2A, SAD participants had significantly higher negative learning rates than HC ($M(\pm STE)$: SAD 0.27(±0.05), HC 0.06(±0.04)), $t(54) = 3.49$, $p = 0.001$, $d = 0.93$, but the two groups showed no differences on positive learning rates ($M(\pm STE)$: SAD 0.22(±0.07), HC 0.34(±0.06)), ps > .15. These findings demonstrate that SAD participants incorporated negative feedback more quickly into their feelings about the self than did HCs. Further, the SAD group did not show a significant difference between positive relative to negative affective updating, $t(20) = −1.07$, $p = 0.29$, in contrast to the strong positivity bias that characterized affective updating in HCs, $t(34) = 4.40$, $p < 0.001$. Taken together, these findings demonstrate that group differences in affective updating were specifically in response to negative (but not positive) feedback, and specific to feelings about the self (but not feelings about others).

Short- and Long-Term Influences of Judges' Feedback on Performance Self-Ratings

Using a mixed-effects ANOVA, we next investigated how participants changed the perception of their own speech performance over time as a function of the difference between the judges' feedback and their own ratings of their performance. Overall,

participants altered their performance self-ratings in the direction of the judges' feedback, $F(1,31) = 52.67$, $p < 0.001$, partial $\eta^2 = 0.63$, though this effect became weaker over time: T2 *versus* T3, $F(1,31) = 9.04$, $p = 0.005$, partial $\eta^2 = 0.23$. Most importantly, in line with our hypothesis, there was a significant feedback valence by group interaction (across both T2 and T3), $F(1,31) = 10.83$, $p = 0.002$, partial $\eta^2 = 0.26$, but no main effect of group on learning (see Figure 2B). This valence by group interaction (across both T2 and T3) was driven by significantly more positive adjustments in the HC compared to the SAD group, $t(31) = 3.05$, $p = 0.005$, and a trend for more negative adjustments in the SAD compared to the HC group, $t(31) = 1.79$, $p = 0.084$. Further, the HC adjusted significantly more to positive than negative feedback mismatch, in line with a positivity bias, $t(18) = 2.47$, $p = 0.024$. The SAD showed the opposite effect, i.e., greater adjustments to negative compared to positive feedback mismatch, $t(13) = 2.45$, $p = 0.029$. No significant valence by group by time interaction ($F(1,31) > 1$) was found, indicating that this interaction was stable over time.

To specifically evaluate whether these effects remained significant at each time point separately (see Figure 2B), planned comparisons revealed that from T1 to T2, SAD participants dynamically adjusted their performance ratings less towards T1 positive feedback than HCs ($M(\pm STE)$: SAD 0.18($\pm$0.05), HC 0.31($\pm$0.04)), $t(31) = 2.07$, $p = 0.047$, and trended towards adjusting their performance ratings more towards T1 negative feedback, relative to HC ($M(\pm STE)$: SAD 0.37($\pm$), HC 0.17($\pm$0.07)), $t(31) = 1.86$, $p = 0.071$ (see Figure 2B). At T3, the group difference in learning from positive social feedback remained significant ($M(\pm STE)$: SAD –0.01($\pm$0.09), HC 0.25($\pm$0.08)), $t(31) = 2.32$, $p = 0.027$, whereas the group difference in learning from negative social feedback became nonsignificant ($M(\pm STE)$: SAD 0.08($\pm$0.09), HC 0.03($\pm$0.08)). There was no significant interaction effect of sex with valence ($F < 1$, see *Supplemental Materials*, Figure S5). In sum, these results demonstrate that social feedback enduringly impacts performance self-ratings in a manner that is positively biased in HC but not in SAD.

Relationship with Questionnaire Measures and Spontaneous Thought Content

We investigated whether positive biases in affective updating were related to symptom severity and individual differences in anxiety. Even when controlling for group status (HC versus SAD) and sex, more anxious participants showed higher negative versus positive learning rates for feelings about the self ($\alpha_{Self\_diff} = \alpha_{SelfPos} - \alpha_{SelfNeg}$) (STAI: partial $r = -0.46$, $p = 0.001$; AAQ-II: $r = -0.29$, $p = 0.037$; FNE: $r = -0.48$, $p < 0.001$). Other anxiety measures showed trends in the same direction (SIAS: partial $r = -0.22$, $p = 0.11$, NSPS: partial $r = -0.19$, $p = 0.16$). Rumination (RRQ: partial $r = -0.37$, $p = 0.006$) and depressive symptoms (PHQ-9: $r = -0.24$, $p = 0.076$) were also negatively related to α_{Self_diff}. We also assessed the relationship between learning rates and trait self-compassion, due to the previous finding that lower self-compassion predicted greater fear of evaluation and severity of SAD (Werner et al., 2012). At present, greater self-compassion was positively related to α_{Self_diff} (SCS: partial $r = 0.39$, $p = 0.003$), indicating that a more compassionate relationship to the self and less clinical symptoms were related to more positive and less negative updating of feelings towards the self.

Finally, we tested whether the difference between positive and negative learning rates for feelings about the self (a_{Self_diff}) during the feedback task was predictive of participants' spontaneous thoughts following the feedback task (i.e., post feedback). As predicted, a_{Self_diff} was negatively correlated with the percentage of negative-affect- and anxiety-related words in the spontaneous thought content following social feedback, even when controlling for spontaneous thought content before receiving feedback (partial robust $r = -0.37$, $p = 0.005$ and $r = -0.30$, $p = 0.028$, respectively, see *Supplemental Materials*, Figure S6). Thus, the more participants dynamically updated their feelings towards the self based on negative relative to positive feedback, the more their spontaneous thoughts after the feedback task contained negative affect- and anxiety-related words (and vice versa). Importantly, the difference between positive and negative learning rates for feelings towards the judges (a_{Judges_diff}) was *not* associated with negative-affect- or anxiety-related thought content after the feedback task (p's > 0.50). Thus, only updating feelings towards the self, but not towards others, related to subsequent negative thought content.

Discussion

This study developed and tested a novel social feedback paradigm to assess the learning mechanisms that generate and maintain the typical positivity or optimism bias towards the self among healthy adults (e.g., Korn et al., 2012; Taylor & Brown, 1988; Weinstein, 1980), and to elucidate how this bias may be altered in SAD—a clinical population in which beliefs and feelings about the social self are generally negative (Clark & Wells, 1995; Hofmann, 2007; Moscovitch, 2009; Rapee & Heimberg, 1997). Specifically, we employed a computational modeling approach to evaluate the dynamic unfolding of how social feedback influences self-directed feelings and self-evaluation. Our findings supported one of our two competing hypotheses: that adults with SAD would update their self-directed feelings and self-perception to a greater extent in response to negative than to positive performance feedback, whereas healthy adults (i.e., without SAD or other psychological disorders) would show the opposite updating bias. Thus, the alternative hypothesis—that adults with SAD would simply learn less from social feedback more generally—was not supported. This study is the first to demonstrate that the positivity bias in the influence of social feedback on self-perception—previously observed in healthy adults (Korn et al., 2012)—is absent or reversed in a clinical sample of socially anxious adults. It is also the first study we know of to apply computational modeling to understanding how feelings about the self—or state *self-esteem*—are dynamically updated or learned in response to social performance feedback (in any population) (for related computational models of social reinforcement learning and state happiness, see e.g., Eldar & Niv, 2015; Eldar, Rutledge, Dolan, & Niv, 2016; Jones et al., 2011; Rutledge, Skandali, Dayan, & Dolan, 2014; Zaki, Kallman, Wimmer, Ochsner, & Shohamy, 2016).

This computational modeling approach advances the field in multiple ways: First and foremost, our novel social learning approach reveals the dynamics of learning from social feedback in SAD and in healthy adults, thereby providing insight into the mechanisms that maintain SAD and underlie perceptions and feelings about the self more generally. Our approach shows how negatively biased learning from social feedback can compound over time, plausibly resulting in a vicious cycle of negative self-evaluation and negative feelings

towards the self that maintains and likely progresses the disorder. Important prior work posits that adults with SAD either notice more negative than positive social cues in the environment, show reduced attention to external social cues more generally, or negatively interpret ambiguous cues (Gilboa-Schechtman, Foa, & Amir, 1999; MacLeod, Mathews, & Tata, 1986; reviewed by Schultz & Heimberg, 2008). By giving clear social feedback in each trial and illustrating visually whether it was more positive or negative than the participant's own self-rating, the current study minimized both the ambiguity of the feedback and the possibility that the SAD group failed to notice positive feedback. The fact that adults with SAD still showed negatively biased, but not overall reduced social learning about themselves and their performance, suggests that *learning* biases may occur independently of or synergistically with known *attentional* biases (Morrison & Heimberg, 2013)—a possibility that future studies could test directly. Thus, this study provides initial evidence that *negatively biased social learning about the self* characterizes and potentially serves to maintain SAD.

More specifically, using an affective updating model, this study demonstrates that social feedback impacts how people feel about themselves and the judges, as a function of learning rates and previous feeling states. This finding is important for a growing literature that suggests strong links between performance monitoring (i.e., error and feedback monitoring) and emotion (for review see Inzlicht, Bartholow, & Hirsh, 2015; Koban, Corradi-Dell'Acqua, & Vuilleumier, 2013; Koban & Pourtois, 2014; Moser, Moran, Schroder, Donnellan, & Yeung, 2013; Proudfit, Inzlicht, & Mennin, 2013). Previous studies have mainly focused on the emotional consequences of negative phenomenon such as errors and negative feedback, and how they are altered in anxious individuals (Aarts & Pourtois, 2012; Koban & Pourtois, 2014; Proudfit et al., 2013). The present findings, in contrast, reveal that one major difference between adults with and without SAD is the extent to which they update how they feel about themselves based on *positive* feedback from others. Adults in the healthy control group showed an even stronger impact of positive feedback than negative feedback on how they felt about themselves, whereas adults with SAD showed less positively biased affective updating. This finding is consistent with growing evidence that SAD is characterized by diminished positive affect in addition to greater negative affect (Amir, Prouvost, & Kuckertz, 2012; Gilboa-Schechtman, Shachar, & Sahar, 2014; Kashdan, 2007). It is also in line with previous findings suggesting that SAD is associated with fear of *positive*, in addition to fear of *negative* social evaluation (Alden, Taylor, Mellings, & Laposa, 2008; Arkin & Appelman, 1983; Sapach, Carleton, Mulvogue, Weeks, & Heimberg, 2015; Wallace & Alden, 1997; Weeks, 2015; Weeks & Howell, 2012).

In addition to influencing how people feel, social feedback provided by 'expert judges' also influenced performance self-ratings, consistent with a growing literature on social influences on affective states (Klucharev, Hytönen, Rijpkema, Smidts, & Fernández, 2009; Koban & Wager, 2016; Zaki, Schirmer, & Mitchell, 2011) and self-perception (Korn et al., 2012). In line with our hypothesis, we again found a positivity bias in healthy participants: When provided with feedback by 'expert judges', participants adjusted their performance self-evaluation to a greater degree towards more positive than negative feedback, consistent with previous findings that suggest a positive learning bias for self-related information in healthy individuals (Sharot et al., 2011; Taylor & Brown, 1988). Importantly, this learning bias

endured for ~1 year following receipt of the judges' feedback. Thus, this positivity bias for the influence of social feedback on ratings of one's performance is long-lasting in healthy individuals, yet enduringly absent, if not reversed, in socially anxious individuals. This group difference is even more striking given that the higher positive self-evaluation of healthy control participants left less room for increases than did the more negative self-evaluation of SAD participants, ruling out the possibility that the observed group difference could be attributed to a regression to the mean.

Implications for Understanding SAD from a Learning Perspective

Only a few studies have attempted to model and predict dynamic changes in emotional states based on computational learning models (Chang, Smith, Dufwenberg, & Sanfey, 2011; Rutledge, Skandali, Dayan, & Dolan, 2014; Xiang, Lohrenz, & Montague, 2013). In clinical populations, the emerging field of computational psychiatry (Montague, Dolan, Friston, & Dayan, 2012; Wang & Krystal, 2014) has focused mainly on decision-making tasks (e.g. Browning, Behrens, Jocham, O'Reilly, & Bishop, 2015) rather than on affective modeling (but see Eldar & Niv, 2015; Eldar, Rutledge, Dolan, & Niv, 2016). The present study suggests that focusing on how emotional states are dynamically altered by social feedback provides a promising avenue to understanding SAD and possibly other clinical disorders from a novel perspective. The stronger learning from negative compared to positive social feedback in SAD may cause the under-estimation of performance in public settings and social situations that is typically observed in this disorder (Hirsch et al., 2003; Rapee & Lim, 1992; Stopa & Clark, 1993). The absent positivity bias in affective updating among adults with SAD provides a mechanistic and computational explanation of the negative emotional relationship to the self (Clark, 2005; Hirsch et al., 2003) and the characteristic low self-esteem and low self-compassion observed in this disorder (Werner et al., 2012). They are consistent with recent work suggesting that socially anxious individuals discount positive social outcomes (Alden et al., 2008; Amir et al., 2012; Weeks, Jakatdar, & Heimberg, 2010) and have reduced memory for positive compared to negative social feedback (Cody & Teachman, 2010; Gilboa-Schechtman et al., 2014). A recent study (Caouette & Guyer, 2016) similarly suggests that more socially anxious participants show reduced positive affect following social acceptance, but increased negative affect in response to social rejection. Finally, our findings support theoretical and empirical work showing that negative views of the social self and processes to manage such views are central to the pathogenesis of SAD (Moscovitch, 2009).

Implications for Understanding Perception and Feelings About the Self Among Healthy Adults

Positive social feedback led to greater learning in non-anxious, healthy adults, in line with a general motivation toward self-enhancement found in healthy people (Brown, 1986). Enhanced learning from positive experiences and social feedback can cause and maintain an overall positive self-perception and potentially even an overestimation of performance. Thus, although a positively biased self-image can be adaptive (Taylor, 1989), evidence suggests that an overly positive and optimistic self-perception can also have negative consequences (e.g. Robins & Beer, 2001). However, given that healthy controls in the present study still learned from negative feedback, albeit to a lesser degree, the positively biased learning

observed among such individuals may be optimal for both a "healthy optimism" and at the same time allow for a flexible and relatively realistic self-image. Future research could employ computational modeling to identify which degree or form of positivity bias is optimal for a healthy and optimistic but also relatively realistic self-image (see also Schneider, 2001; Sharot et al., 2011).

The sociometer theory of self-esteem (Leary et al., 1995), which suggests that self-esteem (i.e. the affective component of the self, Brown, 1993) acts as an indicator of how socially included or rejected an individual feels, helps to account for the current effects of social feedback on affective updating. A recent study showed that healthy adults reported lower state self-esteem when they received negative compared to positive feedback regarding their personality from a confederate (Eisenberger et al., 2011). Using a novel affective updating model, our results are consistent with the sociometer theory of self-esteem (Leary et al., 1995), but they also go beyond these previous findings by demonstrating how *feeling about the self* is updated *asymmetrically* in healthy people, who are more reactive to positive compared to negative affective prediction errors.

How people felt about themselves in response to social feedback impacted their spontaneous self-generated cognition (Cacioppo, Von Hippel, & Ernst, 1997), thereby demonstrating an important convergence between self-reported feeling states and behavioral measures of thought content. Learning rates in affective updating were predictive of subsequent spontaneous negative and anxiety-related thought content across both groups. Importantly, recent work shows that spontaneous thoughts with social and negative emotion-related content were positively correlated with cortisol and alpha-amylase levels both at baseline and following a social stress test that resembled the speech portion of the current study (Engert, Smallwood, & Singer, 2014). Future work could synergize these findings by investigating the relationship among biased affective learning rates, spontaneous thoughts, and stress-related reactivity.

Limitations and Future Directions

Since we only compared clinically diagnosed adults with SAD to low-anxiety, healthy control adults, we cannot make claims regarding the specificity of the present findings for SAD *versus* their generalizability across a wider range of psychological disorders. Our results showed that several questionnaire measures related to anxiety and to ruminative and depressive tendencies were associated with a reduced positivity bias, which thus may be characteristic of anxiety and affective disorders more broadly. In line with this idea, a recent study (Korn, Sharot, Walter, Heekeren, & Dolan, 2014) investigated optimism biases regarding future scenarios in depressed and healthy participants, and showed that major depressive disorder was characterized by the absence of the optimism bias that characterized prospective thinking of healthy individuals. We speculate that the absence of optimism/ positivity bias may be found across different emotional disorders with internalizing aspects, but that the *content* of this learning deficit may be disorder-specific (i.e., concerns about social competence, appearance, and visible signs of anxiety in SAD versus more general self- and future-oriented biases in MDD).

Further, even though we controlled statistically for regression to the mean in our analysis of the changes in performance self-ratings, we cannot completely rule out some of its residual effects. However, the interaction between group and valence cannot be explained by regression to the mean, as this would result in greater learning from more positive feedback in SAD (who had more negative initial self-evaluations than healthy controls) and greater negative learning among healthy controls—exactly the opposite of what we found. The fact that the observed influence of the judges' feedback on self-evaluation decreased over time (from T2 to T3) is also a strong indication of learning effects rather than regression to the mean. In this study, actual judges' ratings (albeit selected) were used to make the task more realistic and to increase external validity compared to randomly generated or standardized feedback—a strength of this study. However, future studies could employ computer-generated ratings to further increase experimental control. In addition, the follow-up time period (T3) was non-standardized (ranging between 7 and 16 months) due to our contacting previous participants at once, and should be standardized in future studies.

Finally, the present sample was not perfectly matched with regard to demographics and only included two male participants with diagnosed SAD, due to difficulties in recruiting male SAD participants that matched our inclusion criteria (e.g., no heavy or daily substance use, which can influence learning processes). We controlled statistically for the potential influence of sex to exclude the possibility that our effects were driven by differences between male and female participants, none of which approached significance, and compared female-only participants across both groups, which demonstrated the same findings, together suggesting that gender did not influence group differences in learning rates. Nonetheless, future studies should aim to assess for potential sex differences in better matched groups.

Conclusions

In sum, we developed and tested a novel social feedback task and computational modeling approach to understand how self-perception is updated in response to social feedback in adults with SAD and healthy controls. Our results demonstrated that healthy adults show a positivity bias when learning from social feedback, both regarding their performance abilities and their feelings towards themselves. This positivity bias was absent, if not reversed, in SAD, thereby providing a dynamic computational mechanism for the consistently negative self-image and low self-esteem in this impairing psychological disorder. Future studies should assess intervention strategies that target and modulate biased social learning, with potential implications for advancing the understanding and treatment of SAD. In particular, it would be of interest to investigate whether negatively biased learning may be counteracted via interventions that boost positive affect (e.g., Alden & Trew, 2013) or kindness towards the self (e.g., self-compassion; Arch et al., 2014). Moreover, research should attempt to ascertain the extent to which socially anxious patients' social learning biases may be differentially impacted by contemporary CBT interventions for SAD such as video feedback (e.g., Orr & Moscovitch, 2010), imagery rescripting (Reimer & Moscovitch, 2015), or behavioral experiments (e.g., Rapee, Gaston, & Abbott, 2009), and by interventions such as cognitive restructuring, cognitive defusion, or mindfulness—which are

specifically designed to provide experiential feedback to facilitate positive reappraisal or reduced impact of the negative self-image which lies at the core of the problem.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Aarts K, Pourtois G. Anxiety disrupts the evaluative component of performance monitoring: An ERP study. Neuropsychologia. 2012; 50(7):1286–1296. DOI: 10.1016/j.neuropsychologia.2012.02.012 [PubMed: 22374184]

Alden LE, Taylor CT, Mellings TMJB, Laposa JM. Social anxiety and the interpretation of positive social events. Journal of Anxiety Disorders. 2008; 22(4):577–590. [PubMed: 17587542]

Alden LE, Trew JL. If it makes you happy: Engaging in kind acts increases positive affect in socially anxious individuals. Emotion. 2013; 13(1):64. [PubMed: 22642341]

Amir N, Prouvost C, Kuckertz JM. Lack of a Benign Interpretation Bias in Social Anxiety Disorder. Cognitive Behaviour Therapy. 2012; 41(2):119–129. DOI: 10.1080/16506073.2012.662655 [PubMed: 22545788]

Arch JJ, Brown KW, Dean DJ, Landy LN, Brown KD, Laudenslager ML. Self-compassion training modulates alpha-amylase, heart rate variability, and subjective responses to social evaluative threat in women. Psychoneuroendocrinology. 2014; 42:49–58. [PubMed: 24636501]

Arkin RM, Appelman AJ. Social anxiety and receptivity to interpersonal evaluation. Motivation and Emotion. 1983; 7(1):11–18.

Bond FW, Hayes SC, Baer RA, Carpenter KM, Guenole N, Orcutt HK, Zettle RD. Preliminary psychometric properties of the Acceptance and Action Questionnaire–II: A revised measure of psychological inflexibility and experiential avoidance. Behavior Therapy. 2011; 42(4):676–688. [PubMed: 22035996]

Brown JD. Evaluations of Self and Others: Self-Enhancement Biases in Social Judgments. Social Cognition. 1986; 4(4):353–376. DOI: 10.1521/soco.1986.4.4.353

Brown JD. Self-esteem and self-evaluation: Feeling is believing. 1993

Brown JD. High self-esteem buffers negative feedback: Once more with feeling. Cognition and Emotion. 2010; 24(8):1389–1404. DOI: 10.1080/02699930903504405

Browning M, Behrens TE, Jocham G, O'Reilly JX, Bishop SJ. Anxious individuals have difficulty learning the causal statistics of aversive environments. Nature Neuroscience. 2015; 18(4):590–596. DOI: 10.1038/nn.3961 [PubMed: 25730669]

Cacioppo JT, Von Hippel W, Ernst JM. Mapping cognitive structures and processes through verbal content: the thought-listing technique. Journal of Consulting and Clinical Psychology. 1997; 65(6): 928. [PubMed: 9420354]

Caouette JD, Guyer AE. Cognitive distortions mediate depression and affective response to social acceptance and rejection. Journal of Affective Disorders. 2016; 190:792–799. DOI: 10.1016/j.jad. 2015.11.015 [PubMed: 26625091]

Chang LJ, Smith A, Dufwenberg M, Sanfey AG. Triangulating the neural, psychological, and economic bases of guilt aversion. Neuron. 2011; 70(3):560–572. DOI: 10.1016/j.neuron. 2011.02.056 [PubMed: 21555080]

Clark DM. A cognitive perspective on social phobia. The essential handbook of social anxiety for clinicians. 2005:193–218.

Clark DM, Wells A. A cognitive model of social phobia. Social phobia: Diagnosis, assessment, and treatment. 1995; 41(68):22–23.

Cody MW, Teachman BA. Post-event processing and memory bias for performance feedback in social anxiety. Journal of Anxiety Disorders. 2010; 24(5):468–479. [PubMed: 20399601]

Eisenberger NI, Inagaki TK, Muscatell KA, Haltom KEB, Leary MR. The neural sociometer: brain mechanisms underlying state self-esteem. Journal of Cognitive Neuroscience. 2011; 23(11):3448–3455. [PubMed: 21452934]

Eldar E, Niv Y. Interaction between emotional state and learning underlies mood instability. Nature communications. 2015; 6

Eldar E, Rutledge RB, Dolan RJ, Niv Y. Mood as Representation of Momentum. Trends in Cognitive Sciences. 2016; 20(1):15–24. [PubMed: 26545853]

Engert V, Smallwood J, Singer T. Mind your thoughts: Associations between self-generated thoughts and stress-induced and baseline levels of cortisol and alpha-amylase. Biological Psychology. 2014; 103:283–291. DOI: 10.1016/j.biopsycho.2014.10.004 [PubMed: 25457636]

Garrett N, Sharot T, Faulkner P, Korn CW, Roiser JP, Dolan RJ. Losing the rose tinted glasses: neural substrates of unbiased belief updating in depression. Frontiers in human neuroscience. 2014; 8doi: 10.3389/fnhum.2014.00639

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

Gilboa-Schechtman, E., Shachar, I., Sahar, Y. Positivity Impairment as a Broad-Based Feature of Social Anxiety. In: Weeks, JW., editor. The Wiley Blackwell Handbook of Social Anxiety Disorder. John Wiley & Sons, Ltd; 2014. p. 409-432.

Hirsch CR, Clark DM, Mathews A, Williams R. Self-images play a causal role in social phobia. Behaviour Research and Therapy. 2003; 41(8):909–921. [PubMed: 12880646]

Hofmann SG. Cognitive factors that maintain social anxiety disorder: A comprehensive model and its treatment implications. Cognitive Behaviour Therapy. 2007; 36(4):193–209. [PubMed: 18049945]

Inzlicht M, Bartholow BD, Hirsh JB. Emotional foundations of cognitive control. Trends in Cognitive Sciences. 2015; 19(3):126–132. DOI: 10.1016/j.tics.2015.01.004 [PubMed: 25659515]

Jones RM, Somerville LH, Li J, Ruberry EJ, Libby V, Glover G, Casey BJ. Behavioral and neural properties of social reinforcement learning. The Journal of Neuroscience. 2011; 31(37):13039–13045. [PubMed: 21917787]

Kashdan TB. Social anxiety spectrum and diminished positive experiences: Theoretical synthesis and meta-analysis. Clinical Psychology Review. 2007; 27(3):348–365. [PubMed: 17222490]

Kessler RC, Chiu WT, Demler O, Walters EE. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. Archives of General Psychiatry. 2005; 62(6):617–627. [PubMed: 15939839]

Kirschbaum C, Pirke KM, Hellhammer DH. The 'Trier Social Stress Test'–a tool for investigating psychobiological stress responses in a laboratory setting. Neuropsychobiology. 1993; 28(1–2):76–81. doi:119004. [PubMed: 8255414]

Klucharev V, Hytönen K, Rijpkema M, Smidts A, Fernández G. Reinforcement learning signal predicts social conformity. Neuron. 2009; 61(1):140–151. [PubMed: 19146819]

Koban L, Corradi-Dell'Acqua C, Vuilleumier P. Integration of Error Agency and Representation of Others' Pain in the Anterior Insula. Journal of Cognitive Neuroscience. 2013; 25(2):258–272. [PubMed: 23163419]

Koban L, Pourtois G. Brain systems underlying the affective and social monitoring of actions: an integrative review. Neuroscience & Biobehavioral Reviews. 2014

Koban L, Wager TD. Beyond conformity: Social influences on pain report and physiology. Emotion. 2016; 16(1):24. [PubMed: 26322566]

Korn CW, Prehn K, Park SQ, Walter H, Heekeren HR. Positively biased processing of self-relevant social feedback. The Journal of Neuroscience. 2012; 32(47):16832–16844. [PubMed: 23175836]

Korn CW, Sharot T, Walter H, Heekeren HR, Dolan RJ. Depression is related to an absence of optimistically biased belief updating about future life events. Psychological Medicine. 2014; 44(3): 579–592. DOI: 10.1017/S0033291713001074 [PubMed: 23672737]

Leary MR, Tambor ES, Terdal SK, Downs DL. Self-esteem as an interpersonal monitor: The sociometer hypothesis. Journal of Personality and Social Psychology. 1995; 68(3)

MacLeod C, Mathews A, Tata P. Attentional bias in emotional disorders. Journal of Abnormal Psychology. 1986; 95(1):15–20. [PubMed: 3700842]

Mattick RP, Clarke JC. Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. Behaviour Research and Therapy. 1998; 36(4):455–470. [PubMed: 9670605]

Montague PR, Dolan RJ, Friston KJ, Dayan P. Computational psychiatry. Trends in Cognitive Sciences. 2012; 16(1):72–80. DOI: 10.1016/j.tics.2011.11.018 [PubMed: 22177032]

Morrison AS, Heimberg RG. Social anxiety and social anxiety disorder. Annual review of clinical psychology. 2013; 9:249–274. DOI: 10.1146/annurev-clinpsy-050212-185631

Moscovitch DA. What Is the Core Fear in Social Phobia? A New Model to Facilitate Individualized Case Conceptualization and Treatment. Cognitive and Behavioral Practice. 2009; 16(2):123–134. DOI: 10.1016/j.cbpra.2008.04.002

Moscovitch DA, Huyder V. The negative self-portrayal scale: development, validation, and application to social anxiety. Behavior Therapy. 2011; 42(2):183–196. [PubMed: 21496505]

Moscovitch DA, Rowa K, Paulitzki JR, Ierullo MD, Chiang B, Antony MM, McCabe RE. Self-portrayal concerns and their relation to safety behaviors and negative affect in social anxiety disorder. Behaviour Research and Therapy. 2013; 51(8):476–486. [PubMed: 23778055]

Moser JS, Moran TP, Schroder HS, Donnellan MB, Yeung N. On the relationship between anxiety and error monitoring: a meta-analysis and conceptual framework. Frontiers in human neuroscience. 2013; 7doi: 10.3389/fnhum.2013.00466

Neff KD. The development and validation of a scale to measure self-compassion. Self and Identity. 2003; 2(3):223–250. DOI: 10.1080/15298860309027

Norman SB, Hami Cissell S, Means-Christensen AJ, Stein MB. Development and validation of an overall anxiety severity and impairment scale (OASIS). Depression and anxiety. 2006; 23(4):245–249. [PubMed: 16688739]

Orr EMJ, Moscovitch DA. Learning to re-appraise the self during video feedback for social anxiety: Does depth of processing matter? Behaviour Research and Therapy. 2010; 48(8):728–737. [PubMed: 20462568]

Pennebaker, JW., Booth, RJ., Francis, ME. Linguistic inquiry and word count: LIWC [Computer software]. Austin, TX: liwc. net; 2007.

Proudfit GH, Inzlicht M, Mennin DS. Anxiety and error monitoring: the importance of motivation and emotion. Frontiers in human neuroscience. 2013; 7

Rapee RM, Gaston JE, Abbott MJ. Testing the efficacy of theoretically derived improvements in the treatment of social phobia. Journal of Consulting and Clinical Psychology. 2009; 77(2):317. [PubMed: 19309190]

Rapee RM, Heimberg RG. A cognitive-behavioral model of anxiety in social phobia. Behaviour Research and Therapy. 1997; 35(8):741–756. [PubMed: 9256517]

Rapee RM, Lim L. Discrepancy between self-and observer ratings of performance in social phobics. Journal of Abnormal Psychology. 1992; 101(4):728. [PubMed: 1430614]

Reimer SG, Moscovitch DA. The Impact of Imagery Rescripting on Memory Appraisals and Core Beliefs in Social Anxiety Disorder. Behaviour Research and Therapy. 2015

Rescorla RA, Wagner AR. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. Classical conditioning: Current research and theory. 1972

Robins RW, Beer JS. Positive illusions about the self: short-term benefits and long-term costs. Journal of Personality and Social Psychology. 2001; 80(2):340. [PubMed: 11220450]

Roy-Byrne P, Craske MG, Sullivan G, Rose RD, Edlund MJ, Lang AJ, et al. Delivery of evidence-based treatment for multiple anxiety disorders in primary care: a randomized controlled trial. Jama. 2010; 303(19):1921–1928. [PubMed: 20483968]

Rutledge RB, Skandali N, Dayan P, Dolan RJ. A computational and neural model of momentary subjective well-being. Proceedings of the National Academy of Sciences. 2014; 111(33):12252–12257.

Safren SA, Heimberg RG, Brown EJ, Holle C. Quality of life in social phobia. Depression and anxiety. 1996; 4(3):126–133. [PubMed: 9166641]

Sapach MJNT, Carleton RN, Mulvogue MK, Weeks JW, Heimberg RG. Cognitive Constructs and Social Anxiety Disorder: Beyond Fearing Negative Evaluation. Cognitive Behaviour Therapy. 2015; 44(1):63–73. DOI: 10.1080/16506073.2014.961539 [PubMed: 25277488]

Schneider SL. In search of realistic optimism: Meaning, knowledge, and warm fuzziness. American Psychologist. 2001; 56(3):250. [PubMed: 11315251]

Schultz LT, Heimberg RG. Attentional focus in social anxiety disorder: Potential for interactive processes. Clinical Psychology Review. 2008; 28(7):1206–1221. DOI: 10.1016/j.cpr.2008.04.003 [PubMed: 18555570]

Schwarz G. Estimating the Dimension of a Model. The Annals of Statistics. 1978; 6(2):461–464. DOI: 10.1214/aos/1176344136

Sharot T, Korn CW, Dolan RJ. How unrealistic optimism is maintained in the face of reality. Nature Neuroscience. 2011; 14(11):1475–1479. DOI: 10.1038/nn.2949 [PubMed: 21983684]

Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (MINI): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. Journal of clinical psychiatry. 1998; 59:22–33.

Spielberger CD. STAI manual for the state-trait anxiety inventory. Self-Evaluation Questionnaire. 1970:1–24.

Spitzer RL, Kroenke K, Williams JBW, Group, P. H. Q. P. C. S. et al. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Jama. 1999; 282(18):1737–1744. [PubMed: 10568646]

Stein MB, Kean YM. Disability and quality of life in social phobia: epidemiologic findings. American Journal of Psychiatry. 2000; 157(10):1606–1613. [PubMed: 11007714]

Stopa L, Clark DM. Cognitive processes in social phobia. Behaviour Research and Therapy. 1993; 31(3):255–267. DOI: 10.1016/0005-7967(93)90024-O [PubMed: 8476400]

Swann WB, Griffin JJ, Predmore SC, Gaines B. The cognitive–affective crossfire: When self-consistency confronts self-enhancement. Journal of Personality and Social Psychology. 1987; 52(5):881. [PubMed: 3585700]

Taylor, SE. Positive illusions: Creative self-deception and the healthy mind. Vol. xv. New York, NY, US: Basic Books; 1989.

Taylor SE, Brown JD. Illusion and well-being: A social psychological perspective on mental health. Psychological Bulletin. 1988; 103(2):193–210. DOI: 10.1037/0033-2909.103.2.193 [PubMed: 3283814]

Trapnell PD, Campbell JD. Private self-consciousness and the five-factor model of personality: distinguishing rumination from reflection. Journal of Personality and Social Psychology. 1999; 76(2):284. [PubMed: 10074710]

Wallace ST, Alden LE. Social phobia and positive social events: The price of success. Journal of Abnormal Psychology. 1997; 106(3):416. [PubMed: 9241943]

Wang XJ, Krystal JH. Computational psychiatry. Neuron. 2014; 84(3):638–654. DOI: 10.1016/j.neuron.2014.10.018 [PubMed: 25442941]

Watson D, Friend R. Measurement of social-evaluative anxiety. Journal of Consulting and Clinical Psychology. 1969; 33(4):448. [PubMed: 5810590]

Weeks JW. Replication and Extension of a Hierarchical Model of Social Anxiety and Depression: Fear of Positive Evaluation as a Key Unique Factor in Social Anxiety. Cognitive Behaviour Therapy. 2015; 44(2):103–116. [PubMed: 25530031]

Weeks JW, Howell AN. The Bivalent Fear of Evaluation Model of Social Anxiety: Further Integrating Findings on Fears of Positive and Negative Evaluation. Cognitive Behaviour Therapy. 2012; 41(2):83–95. DOI: 10.1080/16506073.2012.661452 [PubMed: 22489647]

Weeks JW, Jakatdar TA, Heimberg RG. Comparing and contrasting fears of positive and negative evaluation as facets of social anxiety. Journal of Social and Clinical Psychology. 2010; 29(1):68–94.

Weinstein ND. Unrealistic optimism about future life events. Journal of Personality and Social Psychology. 1980; 39(5):806.

Werner KH, Jazaieri H, Goldin PR, Ziv M, Heimberg RG, Gross JJ. Self-compassion and social anxiety disorder. Anxiety Stress Coping. 2012; 25(5):543–558. DOI: 10.1080/10615806.2011.608842 [PubMed: 21895450]

Werner KH, Jazaieri H, Goldin PR, Ziv M, Heimberg RG, Gross JJ. Self-compassion and social anxiety disorder. Anxiety, Stress & Coping. 2012; 25(5):543–558.

Wittchen HU, Fuetsch M, Sonntag H, Müller N, Liebowitz M. Disability and quality of life in pure and comorbid social phobia. Findings from a controlled study. European Psychiatry. 2000; 15(1):46–58. [PubMed: 10713802]

Xiang T, Lohrenz T, Montague PR. Computational Substrates of Norms and Their Violations during Social Exchange. The Journal of Neuroscience. 2013; 33(3):1099–1108. DOI: 10.1523/JNEUROSCI.1642-12.2013 [PubMed: 23325247]

Zaki J, Schirmer J, Mitchell JP. Social influence modulates the neural computation of value. Psychological Science. 2011; 22(7):894–900. [PubMed: 21653908]

Zaki J, Kallman S, Wimmer GE, Ochsner K, Shohamy D. Social Cognition as Reinforcement Learning: Feedback Modulates Emotion Inference. Journal of Cognitive Neuroscience. 2016
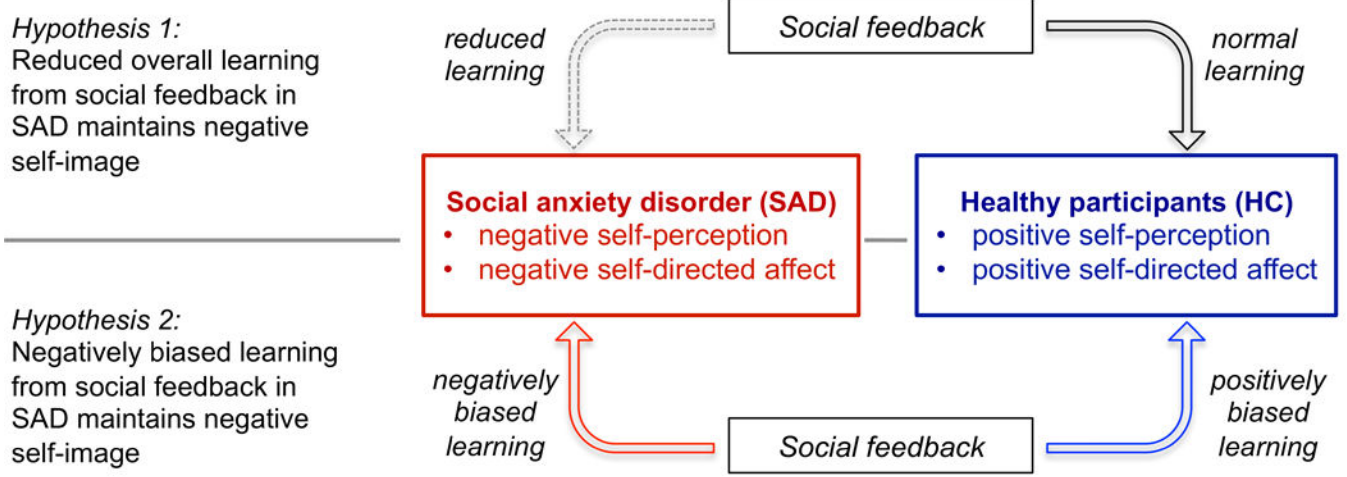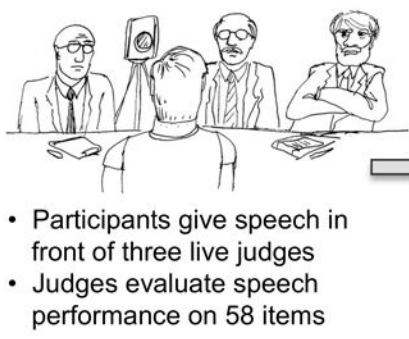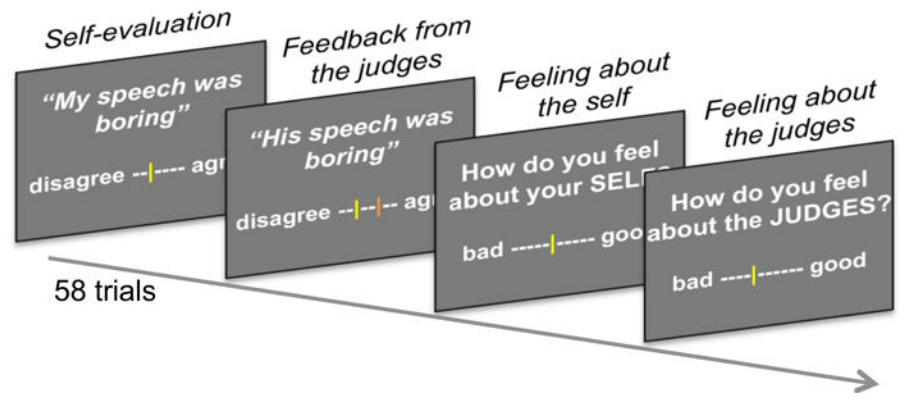
Figure 1. Hypotheses and experimental design

A) Competing hypotheses regarding the maintenance of negative self-perception and negative self-evaluation in social anxiety disorder (SAD) compared to healthy controls (HC). SAD might be characterized by reduced attention to social information and impaired learning from social feedback. Alternatively, SAD might be characterized by negatively biased learning from social feedback, relative to HCs. B) Participants were asked to deliver a speech about their ideal job in front of three live judges, who evaluated their performance on 58 different dimensions relevant to social anxiety (sketch by John Coetzee, retrieved from https://en.wikipedia.org/wiki/Trier_social_stress_test#/media/File:Trier_01.jpg, licensed under a creative commons attribution, CC BY-SA 3.0). C) After the speech, participants performed a self-evaluation and social feedback task. Participants were presented with a rating dimension and asked to evaluate themselves, before viewing the judges' feedback (rating) on this dimension. Participants then rated how they felt about themselves and how they felt about the judges using a VAS with anchors from "very bad" to "very good".
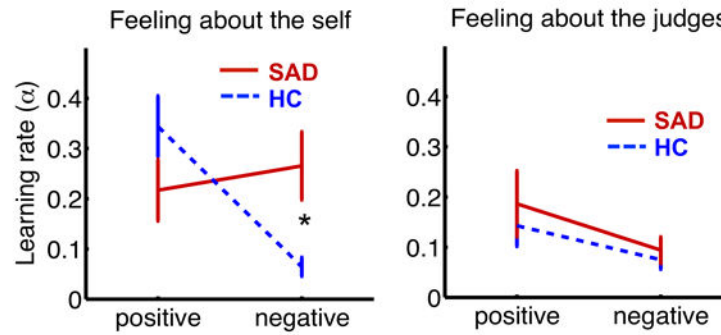
A) Affective learning rates for feelings about the self and the judges

B) Short- and long-term changes in performance self-evaluation
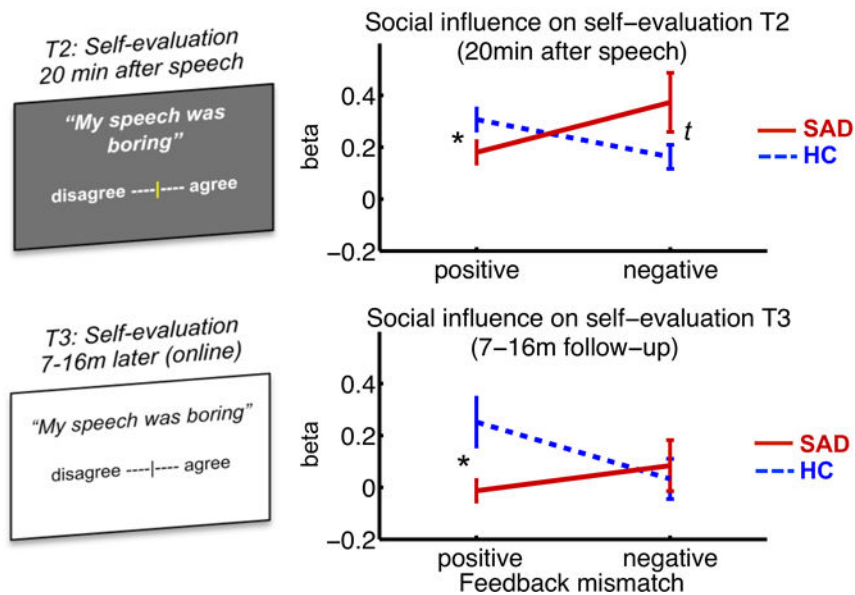

Figure 2. Group differences in affective updating and changes in self-evaluation
A) Learning rates for updating feeling about the self and feeling about the judges, separately modeled for negative and positive affective updating. SAD participants (solid red) updated their feeling about the self more based on negative information than HC, who showed a significant positivity bias (dashed blue line). No group differences were observed for updating feeling about the judges, where both HC and SAD showed stronger positive than negative updating. B) Judges' feedback also differentially impacted self-rated performance on the speech task. The results showed a significant influence of feedback on self-evaluative performance ratings, and a significant valence*group interaction over time: HCs adjusted more towards positive feedback, whereas SAD participants adjusted more towards negative social feedback. Asterisks denote significant group differences; *t* denotes a statistical trend ($p = 0.07$); vertical bars denote standard errors (SEM).

Table 1

Demographic and questionnaire data by group

	Social Anxiety Disorder (SAD) N or Mean (1SD) (n = 21)	Healthy Control (HC) N or Mean (1SD) (n = 35)	p-value
Age	24.0 (4.5)	25.2 (6.1)	0.23
Sex			0.004*
Male	2	17	
Female	19	18	
Race/Ethnicity			0.61
Caucasian	14	24	
Latino/a/Hispanic	2	5	
Asian American	1	2	
African American	1	0	
Biracial	3	2	
Other	0	1	
Education			0.11
High school or less	3	1	
Some college or Associates Degree	9	8	
Bachelor degree	7	14	
Graduate degree	2	11	
Marital status			0.19
Single	12	16	
Married	0	6	
Partnered	8	8	
Dating	1	2	
Other	0	2	
STAI	44.7 (8.7)	29.2 (6.5)	<0.001*
SIAS	33.2 (9.2)	15.6 (3.7)	<0.001*
AAQ-II	35.9 (9.5)	22.1 (7.1)	<0.001*
NSPS	65.6 (15.3)	35.7 (7.0)	<0.001*
SCS	69.5 (18.9)	95.7 (16.0)	<0.001*
RRQ	45.2 (6.7)	30.1 (8.8)	<0.001*
FNE	45.7 (8.1)	27.7 (7.7)	<0.001*
PHQ-9	6.4 (4.5)	2.3 (2.6)	<0.001*
OASIS	10.1 (2.0)	3.9 (2.2)	<0.001*

Note. STAI = State Trait Anxiety Inventory (Spielberger, 1970), SIAS = Social Interaction Anxiety Scale (Mattick & Clarke, 1998), AAQ-II = Acceptance and Action Questionnaire (Bond et al., 2011), NSPS = Negative Self Portrayal Scale (Moscovitch & Huyder, 2011), SCS = Self Compassion Scale (Neff, 2003), RRQ = Rumination and Reflection Questionnaire, FNE = Fear of Negative Evaluation Scale (FNE, Watson & Friend, 1969), PHQ-9 = Patient Health Questionnaire (PHQ-9, Spitzer et al., 1999), OASIS = Overall Anxiety Severity and Impairment Scale (OASIS, Norman et al., 2006). Dimensional data was statistically compared using *t*-tests, categorical data using Pearson Chi-Square tests. Demographic data from 1 HC participant was missing.