Social Surprises and Social Feedback Shape Momentary Mood in Humans

Elena Bagdades

Marjan Biria

Charlotte Burman

Jessica Norman

Raphaelle Delpech

Madeleine Moses-Payne

Naomi Tromans

Lucienne Spencer

Ilina Singh

Eleanor Leigh

Georgina Krebs

Argyris Stringaris

2025-10-07

## Abstract

## Introduction

Humans are inherently social beings, and our daily interactions with others profoundly shape our affect. A single encounter, whether a harsh critique in an interview or an unexpectedly warm conversation, can significantly impact how we feel. Despite the central role social experiences play in our emotional lives, the mechanisms underlying the relationship between these experiences and momentary mood remain vastly underexplored. In this study, we investigate the hypothesis that social surprises, the discrepancy between what we expect from a social interaction and the actual outcome, play a key role in shaping momentary mood. We reverse-translate elements of clinical theory and observation by introducing a novel experimental and computational framework to test this hypothesis in adolescents and adults, including individuals with elevated symptoms of social anxiety and depression.

This centrality of social experiences in shaping our affective lives is rooted in our evolutionary history. Human survival has long depended on our ability to navigate complex social environments. Several evolutionary theories suggest that traits such as cooperation, altruism, and culture emerged because they provided advantages in group living, such as cooperative foraging and shared child-rearing (Tomasello, 2014; Trivers, 1971). In line with this, human emotions are believed to have evolved to help us address social challenges and promote cohesion and cooperation, ultimately enhancing our chances of survival (e.g., Nesse, 1990). Given the role of social behaviour in human evolution, it is unsurprising that social relationships have a profound impact on people’s quality of life and overall mental and physical health (Cacioppo & Cacioppo, 2014; Kuczynski et al., 2020; Wickramaratne et al., 2022). Notably, many psychopathologies are closely tied to interpersonal difficulties, particularly during adolescence (Fett et al., 2015; Lamblin et al., 2017). Impairments in social functioning are especially evident in mood and anxiety disorders, with Social Anxiety Disorder (SAD) being a key example (Saris et al., 2017; Wittchen et al., 2000). Exploring how social interactions shape affect can thus yield valuable insights into human functioning, both in mental health and in disorder.

One promising approach to unpacking this relationship is through computational modelling. Recent advances in this field have begun to uncover the mechanisms of self-reported momentary mood. More specifically, research has highlighted a central role for reward prediction errors (RPEs), in shaping momentary fluctuations in mood (Keren et al., 2021; Rutledge et al., 2014). According to this framework, mood can be explained by integrating over the history of events, especially RPEs, that occur in our environment. Positive RPEs, where outcomes exceed expectations, tend to elevate mood, while negative RPEs, where outcomes fall short, tend to suppress it. For example, an unexpected bonus at work might boost mood, whereas receiving the same bonus when expecting a bigger one, could lead to disappointment and to a lower mood. Computational modelling provides a powerful tool for capturing how past experiences shape current mood, including formalising the extent to which past events continue to exert influence over time. For instance, incorporating a forgetting factor allows models to account for the greater weight often given to more recent experiences; known as the recency bias. Extending this research into the social domain is essential as it will allow us to examine whether similar computational mechanisms underlie mood dynamics across different reward types.

Much of the computational literature on momentary mood has focused on the happy–unhappy spectrum, while other types of mood are largely understudied. Specifically, while the processes involved in pathological anxiety, such as in SAD have been extensively studied, the computations involved in anxious mood remain poorly understood. Clinical theory offers valuable insight here. The notion of belief disconfirmation, closely related to prediction error, is foundational in cognitive therapy for anxiety disorders (Clark & Wells, 1995; Salkovskis et al., 2007). For example, treatment for SAD involves generating positive prediction errors that arise when individuals are asked to track their expectations against the outcomes of social interactions during therapy (Clark, 1999). This therapeutic principle can be ‘reverse-translated’ to guide experimental investigations into the mechanisms of momentary anxiety, particularly within social contexts.

In this article, we address key gaps in the literature by examining the computations underlying momentary happy and anxious mood in the context of social interactions. Specifically, we extend computational models of mood into the social domain, incorporate anxious affect, and investigate how these processes vary across symptom dimensions. Given the heightened salience of social rewards and punishments in adolescents (Foulkes & Blakemore, 2016), as well as the rise in mental health problems related to interpersonal difficultiles in this age group (Lamblin et al., 2017), we recruited adolescents aged 14-18 through schools, as well as adults aged 18-45 through the local community and Prolific. We developed a novel experimental paradigm designed to elicit social PEs in which participants interacted with virtual players and were asked to report their momentary happy and anxious mood in response to social feedback. This task was co-designed with a youth advisory group to ensure clarity and engagement for younger participants. We preregistered the hypothesis (<https://osf.io/73zsg/>) that social PEs would be central to affective responses and, more specifically, would be positively associated with happy mood and negatively with anxious mood. To test this, we fitted and compared a series of hierarchical Bayesian computational models that systematically varied in their inclusion and specification of social PEs, expectations, and outcomes. Crucially, this included a model without social PEs, allowing us to directly evaluate the added explanatory value of social PEs relative to a plausible alternative. Our findings suggest that both happy and anxious mood are shaped by immediate social feedback, as well as by the history positive and negative social PEs. In exploratory analyses, we extended the best-fitting model to assess how social anxiety and depressive symptoms modulated these computational parameters. Our results showed that people with higher social anxiety symptoms exhibit a stronger influence of social feedback on their momentary mood and anxiety. Further, individuals with higher depressive symptoms showed a bigger effect of negative social PEs on their mood. This could suggest that individuals at risk for social anxiety and depression are more susceptible to social appraisals when reporting their momentary affect. These findings offer insight into how affective dynamics may become dysregulated in psychopathology and point to potential early markers of risk.

## Results

### Participants

We analysed data from 185 participants aged 14–45 years, recruited from three different sources: online participants (n = 97), university students (n = 43), and school students (n = 44).

### Experimental setup

Participants first completed questionnaires assessing social anxiety and depressive symptoms. Depending on their age group, these included the Liebowitz Social Anxiety Scale – Self-Report (LSAS; Fresco et al. (2001)), the Centre for Epidemiologic Studies Depression Scale (CES-D; Radloff (1977)), the LSAS for Children and Adolescents – Self-Report (LSAS-CA; Leigh & Clark (2022)), and the Revised Child Anxiety and Depression Scale (RCADS; Chorpita et al. (2000)). They then completed an online task simulating socially evaluative interactions (see Fig.1). They were told that four different “virtual players” would rate them after they described images aloud while being video recorded. Each trial began with the presentation of the virtual player assigned to that trial, followed by the participant’s prediction of how well they expected to be rated. Participants then described the image for 15 seconds and received feedback from the virtual player on a scale from 0 to 100. Afterward, they reported their momentary mood and anxiety. The virtual players varied in their level of criticality, and feedback was manipulated to generate a range of social prediction errors (PEs), including negative (PE−), moderate positive (PE+), strong positive (PE++), and neutral (PEN), based on each player’s typical rating behaviour. Trial order, feedback, and stimuli were randomized.

|  |
| --- |
| Figure 1: Structure of a trial in the experimental task. Participants interacted with four different virtual players throughout the task. During the first four trials, each virtual player was introduced alongside a histogram illustrating their rating tendencies, designed to induce expectations about how critical each player would be. From the fifth trial onward, only the name and picture of the virtual player providing the upcoming rating were shown. After this presentation, participants rated how well they expected to perform on that trial (*Prediction*), then described a picture to the virtual player for 15 seconds while being video recorded. They subsequently received a score out of 100 from the virtual player (*Feedback*) and reported their current levels of happy and anxious mood. |

### Momentary anxiety and mood were best explained by asymmetric PEs and social feedback amount.

Out of the eight computational models compared, the best-fitting one for both mood and anxiety, included an effect of social feedback and asymmetric effects for positive and negative prediction errors (PEs). Crucially, this model outperformed alternatives that excluded PEs, supporting the idea that social PEs contribute to shaping affect, in line with our hypothesis. Social feedback was modelled as the outcome of the current trial, while the PE terms were represented as recency-weighted sums of all prior positive and negative PEs, respectively. Specifically, self-reported mood and anxiety at trial for individual was modelled as

where represents the individual-specific baseline of mood or anxiety, and and capture the sensitivity of individual to the history of positive and negative PEs, respectively. The parameter is a forgetting factor that determines the influence of prior trials on affect. represents the social feedback on trial t, and captures its immediate effect on momentary affect for participant .

|  |
| --- |
| Figure 2: Reported and predicted mood and anxiety across trials for two participants. Each panel shows z-scored values across trials, with mood plotted in the left column and anxiety in the right column. Rows represent individual participants. Reported responses are shown in grey, predicted mood in blue, and predicted anxiety in red. Predicted values are derived from the posterior mean of a hierarchical Bayesian model. |

### Group-Level Parameter Estimates

**Social Feedback Effects ()**  
Social feedback had a strong effect on both mood and anxiety. For momentary mood, social feedback had a substantial positive effect (mean **()** = 0.22, 95% HDI [0.19, 0.26]), indicating that more positive social feedback was associated with increased mood. In contrast, for momentary anxiety, social feedback exerted a negative effect (mean **()** = -0.09, 95% HDI [-0.11, -0.07]), suggesting that more positive social feedback was associated with reduced anxiety.

**Positive Prediction Error Effects ()**  
The effect of positive PEs varied across mood and anxiety models. For momentary mood, the effect was small and not substantial (mean **()** = 0.01, 95% HDI [0.00, 0.03]). However, for momentary anxiety, there was a small but substantial negative effect of positive PEs (mean **()** = -0.04, 95% HDI [-0.06, -0.03]), indicating that experiencing outcomes better than expected was linked to reduced anxiety.

**Negative Prediction Error Effects ()**  
Negative PEs did not show a strong association with momentary mood (mean **()** = -0.01, 95% HDI [-0.02, 0.01]). For momentary anxiety, however, there was a small but substantial positive effect (mean **()** = 0.04, 95% HDI [0.02, 0.06]), suggesting that worse-than-expected outcomes were associated with increased anxiety.

**Forgetting Factor ()**  
The forgetting factor determines how much influence past prediction errors (PEs) have on current affect, with values bounded between 0 and 1. A value of implies full recency, where only the most recent trial influences affect, while implies equal weighting of all past trials, regardless of their temporal distance. The group-level estimate of was 0.93 for both mood (95% HDI [0.91, 0.95]) and anxiety (95% HDI [0.92, 0.95]), indicating that past trials have a fairly long-lasting impact on current affect.

These results partially support our hypothesis that social surprises, operationalized here as PEs, play a key role in shaping momentary mood and anxiety. While the best-fitting model included terms for both positive and negative PEs, and outperformed models without them, only social feedback and not the PE themselves showed a substantial effect on mood at the group level. In contrast, anxiety was modulated by both positive and negative PEs, indicating that momentary anxiety is more sensitive to social surprises than mood in this task context.

### Moderation of effects by social anxiety symptoms

To assess whether social anxiety symptoms (SA), as measured by the LSAS, moderated any of the model parameters, we examined interaction terms from the hierarchical Bayesian models predicting momentary anxiety and mood. For both mood and anxiety, the interaction term for social feedback , , showed the most robust effect. For anxiety, this interaction was negative (mean = –0.04, 95% HDI [–0.06, –0.02], see (**param\_distributions?**)), suggesting that higher SA amplified the influence of social feedback on decreasing anxiety. In contrast, for mood, the interaction was positive (mean = 0.04, 95% HDI [0.00, 0.07], see (**param\_distributions?**)), indicating that SA also strengthened the positive impact of social feedback on mood.

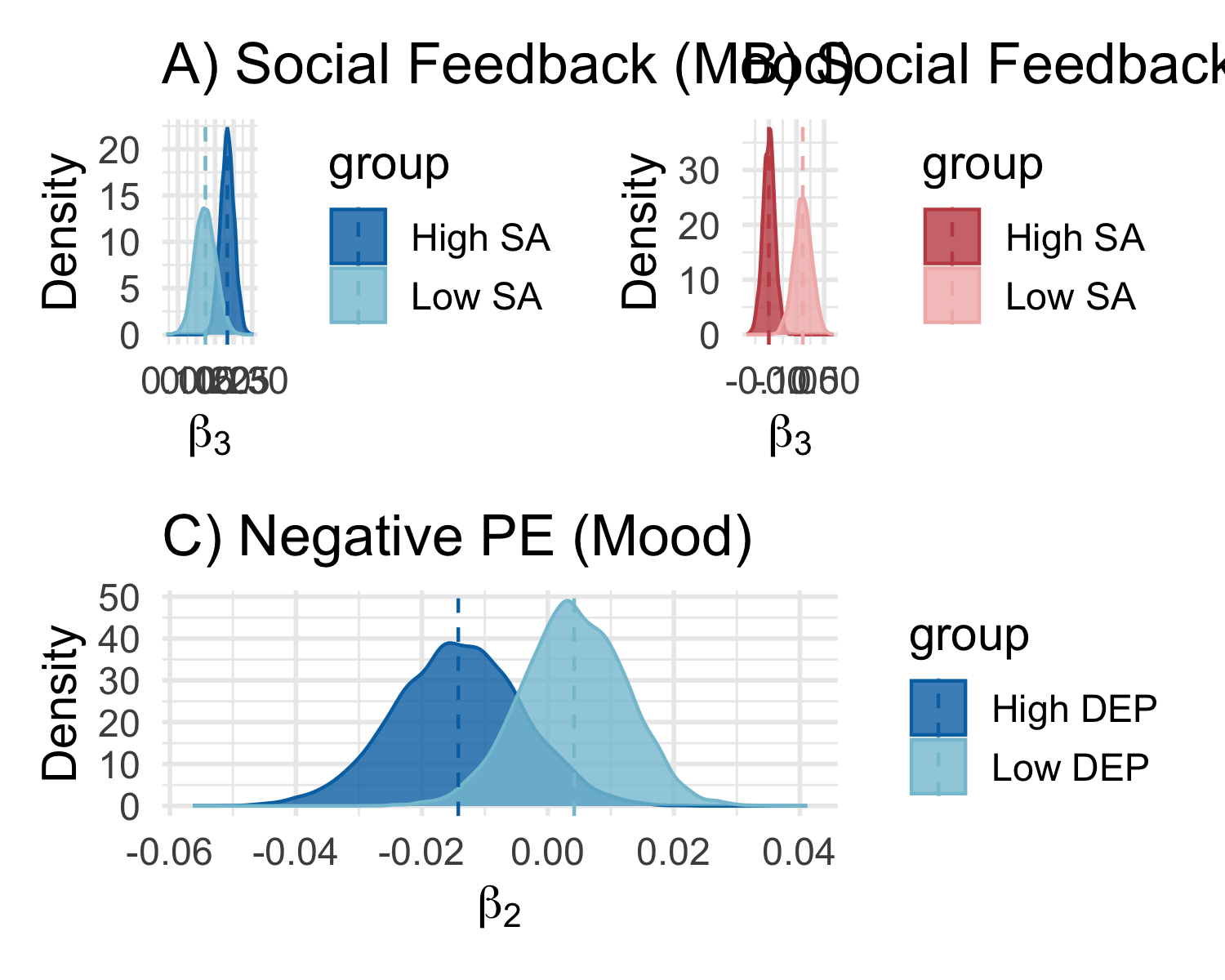
Other interaction terms between SA and the intercept ( ) , positive PEs (), negative PEs (), and the forgetting factor () had credible intervals that overlapped zero for both outcomes, suggesting inconclusive or negligible moderation effects. Similarly, the main effects of SA on momentary anxiety (mean = 0.22, 95% HDI [–0.49, 0.92]) and mood (mean = –0.16, 95% HDI [–0.81, 0.56]) were also uncertain.

To complement the model-based findings, we computed the posterior distribution of Pearson correlations between each participant-level parameter estimates and SA scores. The parameter for social feedback () showed a strong negative correlation with SA in predicting anxiety (mean r = –0.29, 95% CI [–0.35, –0.23], Pr(r < 0) = 100%; see (**posterior\_corr\_distributions?**)), and a moderate positive correlation in predicting mood (mean r = 0.16, 95% CI [0.11, 0.19], Pr(r > 0) = 100%; see (**posterior\_corr\_distributions?**)). These results provide further support for a moderation of the influence of social feedback on momentart affect, by SA. Correlations for all other parameters were weaker and had posterior intervals that included zero.

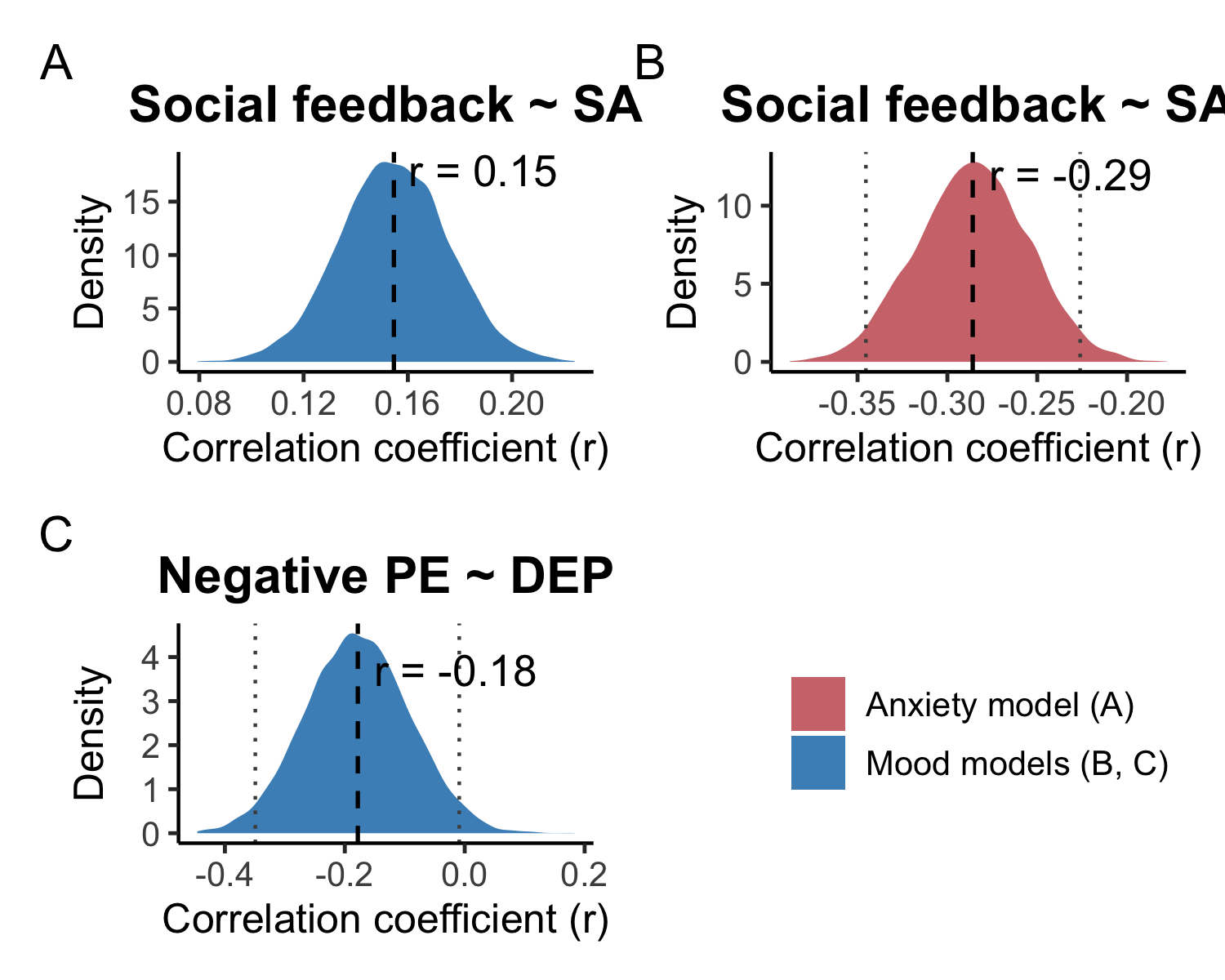
### Moderation of effects by depression symptoms

To assess whether depression symptoms (DEP), as measured by the RCADS in the student sample and the CES-D in all other samples (z-scored within each sample prior to modelling), moderated any of the model parameters, we examined interaction terms from the hierarchical Bayesian models predicting momentary anxiety and mood. For both mood and anxiety models, the interaction terms between DEP and the intercept (), positive PEs (), negative PEs (), social feedback () and the forgetting factor ( )had credible intervals that overlapped zero, suggesting inconclusive or negligible moderation effects. Similarly, the main effects of DEP on momentary anxiety (mean = 0.17, 95% HDI [–0.49, 0.87]) and mood (mean = –0.10, 95% HDI [–0.85, 0.53]) were also uncertain.

We also computed the posterior distribution of Pearson correlations between each participant-level parameter estimates and DEP scores. According to this analysis, the parameter for social feedback ( ) showed a moderate negative correlation with DEP in predicting mood (mean r = –0.18, 95% CI [–0.35, –0.01], Pr(r < 0) = 97.9%). This suggest that individuals with higher DEP, exhibited a stronger negative effect of negative PEs on their mood. Correlations for all other parameters were weaker and had posterior intervals that included zero.



Posterior distributions by SA and DEP groups of A) Social Feedback for mood , B) Social Feedback for anxiety and C) Negative PE for mood.



Posterior correlation distributions for A) Social Feedback and SA for anxiety, B) Social Feedback and SA for mood and C) Negative PE and DEP for mood.

## Discussion

* notes: Interaction effect for Social Feedback x SA can be interpreted as a susceptibility to PE and a target opportunity for treatment.

## **Methods**

### Participants

#### Online Participants

We recruited 106 online participants between the ages of 18-45 using Prolific ([www.prolific.com](https://prolific.com/)). See Table 1 for participants demographic information. Eight participants were excluded due to poor engagement with the task, as evidenced by their video recordings. We required that our participants be adults living in the United Kingdom or the United States, that they speak fluent English, have no cognitive impairment or dementia, and that over 90% of their previous jobs have been approved. We also required that participants had not taken part in any of our pilot studies (which were similar to this one). Participants were paid at a rate of £9/hr as compensation for their time.

#### UCL Community Participants

We recruited 47 participants between the ages of 18-25 from the University College London (UCL) community, using posters and through the UCL psychology subject pool (SONA). Four participants were excluded due to poor engagement with the task, as evidenced by their video recordings. Upon registration, participants were invited to book a testing session, which they were instructed to complete in a quiet and private space. At the time of the testing session, they received a link to the online task. Participants were compensated with a £10 Love2shop or Amazon voucher for their time.

#### School Participants

We recruited 49 participants between the ages of 14-25 through schools (via physical flyers or via school newsletters). Five participants were excluded due to poor engagement with the task, as evidenced by their video recordings. Schools were recruited either by direct contact or by advertising through the Anna Freud school network newsletter. For students under the age of 16, parental opt-out consent was required. Upon registration, participants were invited to book a testing session, which they were instructed to complete in a quiet and private space. At the time of the testing session, they received a link to the online task. Participants were compensated with a £10 Love2shop voucher for their time. As an additional incentive, schools were also offered talks on mental health, careers in psychology or related topics.

Table 1: Demographic Characteristics by Dataset Group

| Dataset Group | N | Age (Mean ± SD) | Female (Sex) | Male (Sex) | Depression | Anxiety |
| --- | --- | --- | --- | --- | --- | --- |
| Prolific participants aged 18–25 | 62 | 22.5 (1.9) | 39 (62.9%) | 23 (37.1%) | 15 (24.2%) | 14 (22.6%) |
| Prolific participants aged 26–45 | 36 | 33.9 (5.9) | 22 (61.1%) | 14 (38.9%) | 6 (16.7%) | 5 (13.9%) |
| Local Community aged 18–25 | 43 | 21.7 (2.5) | 40 (93.0%) | 3 (7.0%) | 2 (4.7%) | 2 (4.7%) |
| School participants aged 14-18 | 44 | 16.4 (1.1) | 27 (61.4%) | 17 (38.6%) | 2 (4.5%) | 6 (13.6%) |

### Questionnaires and Experimental Task

The online questionnaires and experimental task were created and hosted on Gorilla Experiment Builder ([www.gorilla.sc](http://www.gorilla.sc/)). The experiment took approximately 50 minutes to complete. Participants first provided informed consent as approved by UCL’s Research Ethics Committee. Participants then completed a series of questionnaires. First, they completed a survey about their demographics, mental health diagnoses and psychotropic medication use.  Then, participants completed the Mini-Social Phobia Inventory (Mini-SPIN; Connor et al. (2001)), the Generalized Anxiety Disorder 7-item Scale (GAD-7; Spitzer et al. (2006)), and the Affective Reactivity Index (ARI; Stringaris et al. (2012)). School participants additionally completed the Lebowitz Social Anxiety Scale for Children and Adolescents – Self-Report (LSAS-CA; Leigh & Clark (2022)) and the Revised Child Anxiety and Depression Scale (RCADS; Chorpita et al. (2000)). Online and UCL community participants completed the adult version of the Lebowitz Social Anxiety Scale – Self-Report (LSAS; Fresco et al. (2001)), the Centre for Epidemiologic Studies Depression Scale (CES-D; Radloff (1977)), the Adult ADHD Self-Report Scale (ASRS; Kessler et al. (2005)), the Drug Use Disorders Identification Test (DUDIT; Berman et al. (2002)), the Alcohol Use Disorders Identification Test (AUDIT; Saunders et al. (1993)), and the Body Image Questionnaire (BIQ; Cash & Szymanski (1995)). In the present study, we report analyses only from the Mini-SPIN and CES-D and the low mood RCADS subscale.

After completing the mental health questionnaires and passing two attention checks, participants proceeded to an experimental task. In this task, participants were informed that they would practice speaking to others and receive performance ratings from “virtual players” based on how well they came across. The task consisted of 48 trials. On each trial, participants were first shown a picture of the virtual player who would be providing feedback for that trial. For the first trial with each of the four virtual players, the image was accompanied by a histogram depicting that player’s previous ratings of others, in order to establish an initial expectation regarding the player’s level of criticality. The four virtual players varied in their rating style, ranging from very easy-going to highly critical.

After viewing the virtual player, participants were asked to report their expectation regarding how they would be rated on that specific trial, on a scale from 0 to 100. Next, they were shown a picture and instructed to describe it to the virtual player while being video recorded for 15 seconds. Following this, they received feedback from the virtual player in the form of a percentage score. Finally, participants rated their current mood and anxiety levels on a scale from 0 to 100 (with 0 indicating “Very unhappy” or “Very relaxed,” and 100 indicating “Very happy” or “Very nervous/uncomfortable,” respectively).

To induce a variety of social PEs, we manipulated the feedback with the aim of generating two positive (PE+), two bigger positive (PE++), four negative (PE−) and four neutral PEs (PEN; indicating no PE) per virtual player. These were based on the virtual player’s histogram mean. To do this, we used the mean of the virtual players’ histograms (29, 37, 63, 71). PE+ feedback was generated by adding a randomly sampled value (from a normal distribution with mean = 12, SD = 3, range = 12–20) to the histogram mean. PE++ feedback was generated by further adding 10 ±1 to the PE+ value. PE− feedback was generated by subtracting a similarly sampled value from the histogram mean. PEN feedback was set as either the exact histogram mean or one point above or below it. The order of trials with each virtual player, the feedback received, the mood and anxiety rating prompts, and the picture stimuli were all randomized across participants and trials.

This task was developed in collaboration with members of a Young People Advisory Group (YPAG), who provided feedback on multiple pilots. Their input helped refine the phrasing of task instructions and questions to ensure accessibility for young people. Additionally, they evaluated the believability of the virtual players across several iterations of the task. Evidence from piloting suggests the task is socially salient and believable, as indicated by: (1) higher overall anxiety ratings among individuals with elevated social anxiety symptoms, and (2) higher anxiety ratings in pilots that included a video recording component compared to those without it (see Supplement).

### Computational Modelling

Our computational models were chosen based on (1) theoretical considerations and (2) preliminary model comparisons using pilot data (n=37). Eight models were chosen, which were variations of a model developed by Rutledge et al. (2014), each differing on their inclusion and specification of social PEs, expectations, and outcomes. This approach allowed us to identify the combination of social appraisal variables that best explained self-reported momentary mood and anxiety ratings in our task and to isolate the unique contribution of each variable.

#### Calculation of social appraisal variables

Within this approach, the outcome amount, *,* was defined as the percentage score received by the virtual player on trial .

Social PEs were defined as the difference between the actual outcomeand the participant-reported expectation, , on that trial.

In some models, the PE term was divided into separate components for positive and negative social prediction errors, assuming an asymmetric influence on affect.

#### Hierarchical Bayesian Framework

All models were fit within a hierarchical Bayesian framework using the rstan package (Stan Development Team, 2025) in RStudio (2024). In each model, at least one of the social appraisal variables (i.e., social expectations, prediction errors or outcomes) was modelled as a time-decayed accumulator. Specifically, self-reported mood or anxiety at trial for individual was modelled as a normally distributed outcome:

Here, denotes the value of the social appraisal on trial *j.* The parameter is an individual-specific forgetting factor that determines the influence of prior trials on self-reported affect. When, only the most recent trial contributes to affect, indicating no memory of past events. In contrast, reflects full accumulation, where all prior social appraisals are weighted equally.

Model estimation was performed via Hamiltonian Monte Carlo, using four independent Markov chains. Each chain produced 2,000 post-warmup samples after an initial 2,000-iteration warmup, yielding a total of 8,000 samples for posterior inference.

The hierarchical structure of the models relied on partial pooling, where individual-level parameters were assumed to be drawn from group-level normal distributions whose means and standard deviations were estimated from the data. This allowed for individual variation while also leveraging population-level structure. For example, individual parameters were specified as follows:

$$
\begin{align}\beta\_{0i} &= \mu\_{\beta\_0} + \epsilon\_{\beta\_{0i}} \\\beta\_{1i} &= \mu\_{\beta\_1} + \epsilon\_{\beta\_{1i}} \\&\vdots \notag \\\beta\_{ni} &= \mu\_{\beta\_n} + \epsilon\_{\beta\_{ni}} \\\gamma\_i &= \Phi(\mu\_{\gamma} + \epsilon\_{\gamma\_i}) \\\sigma\_i &= \exp(\mu\_{\sigma} + \epsilon\_{\sigma\_i}) \end{align}
$$

where are individual-level deviations. Parameters bounded to specific intervals were sampled in an unconstrained latent space and transformed accordingly. Specifically, the forgetting factor γ, which is bounded between 0 and 1, was transformed using a probit link function to ensure it remained within this interval. Standard deviation parameters were constrained to be positive through exponential transformations. Weakly informative priors were applied to all group-level parameters.

#### Summary of Fitted Models

For our first aim, we fit and compared eight different computational models for mood and anxiety, respectively. The models differed in their inclusion and specification of social appraisals, expectations, PEs and Outcomes. Notably, Models 7 and 8 incorporated asymmetric effects for positive and negative PEs. Models 1 and 2 were modified versions of the standard model proposed by Rutledge et al. (2014) and the primacy model by Keren et al. (2021), respectively, adapted to fit our task-specific data.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 1: Summary table of models compared   | Model | Expectation | Prediction\_Error | Outcome | | --- | --- | --- | --- | | 1 | Recency weighted E | Recency weighted PE | - | | 2 | Primacy weighted E | Recency weighted PE | - | | 3 | Recency weighted E | - | O(t) | | 4 | - | Recency weighted PE | O(t) | | 5 | - | PE(t) | Recency weighted O | | 6 | Recency weighted E | PE(t) | - | | 7 | - | Asymmetric recency weighted PE | Recency weighted O | | 8 | - | Asymmetric recency weighted PE | O(t) | | Note. E = reported expectation; PE = social prediction error; O = Outcome. | | | | |

### Aim 1: Identify the computations underlying momentary mood and anxiety in our social task

#### Model Comparison and Evaluation

To evaluate and compare the performance of computational models, we employed a combination of information-theoretic and predictive accuracy metrics. Specifically, we used the loo package in R to calculate the Leave-One-Out Information Criterion (LOOIC) and the Widely Applicable Information Criterion (WAIC) to assess out-of-sample predictive fit while accounting for model complexity. Both criteria estimate the expected log predictive density for new data and are particularly well-suited for Bayesian models, with lower values indicating better fit. In addition to these criteria, we computed Mean Squared Error (MSE) and Mean Absolute Error (MAE) as measures of predictive accuracy, reflecting the average squared and absolute deviations between predicted and observed values, respectively. Finally, we included the coefficient of determination (R²) as an indicator of explained variance, providing an interpretable summary of model fit. Together, these complementary metrics allowed for a comprehensive assessment of model performance across both explanatory and predictive dimensions. Posterior predictive checks were conducted on the winning model to verify that the fitted model is compatible with our observed data.

The 95% Bayesian Highest Density Interval (HDI) was calculated for the mean of each group-level parameter. A social appraisal parameter was deemed to have a meaningful influence if its 95% HDI did not cross zero, suggesting a statistically credible effect.

### Aim 2: Explore how these computations vary with individual differences in mental health symptoms

To examine individual differences in affective dynamics, we extended the winning model by including a main effect of a participant-level moderator (either social anxiety or depression symptoms) and its interaction with all free parameters. These moderators were included as between-subject predictors that modulate both baseline affect and sensitivity to social signals over time.

Specifically, mood and anxiety were modelled as functions of recency-weighted positive and negative prediction errors, outcome feedback, and a participant-specific moderator Mi​ (e.g., LSAS score for social anxiety, RCADS/CES-D score for depression):

where are individual-level deviations.

to be added:

* (in methods) Paragraph on transparency (will create gorilla links where people can preview experiment ;e.g. <https://app.gorilla.sc/openmaterials/1080493> (will need to publish)

## References

Berman, A. H., Bergman, H., Palmstierna, T., & Schlyter, F. (2002). *Drug use disorders identification test*. <https://doi.org/10.1037/t02890-000>

Cacioppo, J. T., & Cacioppo, S. (2014). Social Relationships and Health: The Toxic Effects of Perceived Social Isolation. *Social and Personality Psychology Compass*, *8*(2), 58–72. <https://doi.org/10.1111/spc3.12087>

Cash, T. F., & Szymanski, M. L. (1995). The development and validation of the body-image ideals questionnaire. *Journal of Personality Assessment*, *64*(3), 466–477. <https://doi.org/10.1207/s15327752jpa6403_6>

Chorpita, B. F., Yim, L., Moffitt, C., Umemoto, L. A., & Francis, S. E. (2000). Assessment of symptoms of DSM-IV anxiety and depression in children: A revised child anxiety and depression scale. *Behaviour Research and Therapy*, *38*(8), 835–855. <https://doi.org/10.1016/S0005-7967(99)00130-8>

Clark, D. M. (1999). Anxiety disorders: why they persist and how to treat them. *Behaviour Research and Therapy*, *37*, S5–S27. <https://doi.org/10.1016/S0005-7967(99)00048-0>

Clark, D. M., & Wells, A. (1995). *A cognitive model of social phobia* (pp. 69–93). The Guilford Press.

Connor, K. M., Kobak, K. A., Churchill, L. E., Katzelnick, D., & Davidson, J. R. T. (2001). Mini-SPIN: A brief screening assessment for generalized social anxiety disorder. *Depression and Anxiety*, *14*(2), 137–140. <https://doi.org/10.1002/da.1055>

Fett, A.-K. J., Shergill, S. S., & Krabbendam, L. (2015). Social neuroscience in psychiatry: unravelling the neural mechanisms of social dysfunction. *Psychological Medicine*, *45*(6), 1145–1165. <https://doi.org/10.1017/S0033291714002487>

Foulkes, L., & Blakemore, S.-J. (2016). Is there heightened sensitivity to social reward in adolescence? *Current Opinion in Neurobiology*, *40*, 81–85. <https://doi.org/10.1016/j.conb.2016.06.016>

Fresco, D. M., Coles, M. E., Heimberg, R. G., Liebowitz, M. R., Hami, S., Stein, M. B., & Goetz, D. (2001). The Liebowitz Social Anxiety Scale: a comparison of the psychometric properties of self-report and clinician-administered formats. *Psychological Medicine*, *31*(6), 1025–1035. <https://doi.org/10.1017/s0033291701004056>

Keren, H., Zheng, C., Jangraw, D. C., Chang, K., Vitale, A., Rutledge, R. B., Pereira, F., Nielson, D. M., & Stringaris, A. (2021). The temporal representation of experience in subjective mood. *eLife*, *10*, e62051. <https://doi.org/10.7554/eLife.62051>

Kessler, R. C., Adler, L., Ames, M., Demler, O., Faraone, S., Hiripi, E., Howes, M. J., Jin, R., Secnik, K., Spencer, T., Ustun, T. B., & Walters, E. E. (2005). The World Health Organization Adult ADHD Self-Report Scale (ASRS): a short screening scale for use in the general population. *Psychological Medicine*, *35*(2), 245–256. <https://doi.org/10.1017/s0033291704002892>

Kuczynski, A. M., Kanter, J. W., & Robinaugh, D. J. (2020). Differential associations between interpersonal variables and quality-of-life in a sample of college students. *Quality of Life Research*, *29*(1), 127–139. <https://doi.org/10.1007/s11136-019-02298-3>

Lamblin, M., Murawski, C., Whittle, S., & Fornito, A. (2017). Social connectedness, mental health and the adolescent brain. *Neuroscience & Biobehavioral Reviews*, *80*, 57–68. <https://doi.org/10.1016/j.neubiorev.2017.05.010>

Leigh, E., & Clark, D. M. (2022). Establishing the psychometric properties of the self-report liebowitz social anxiety scale for children and adolescents in a general population adolescent sample. *European Journal of Psychological Assessment*, *38*(4), 243–250. <https://doi.org/10.1027/1015-5759/a000659>

Nesse, R. M. (1990). Evolutionary explanations of emotions. *Human Nature*, *1*(3), 261–289. <https://doi.org/10.1007/BF02733986>

Radloff, L. S. (1977). The CES-d scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, *1*(3), 385–401. <https://doi.org/10.1177/014662167700100306>

Rutledge, R. B., Skandali, N., Dayan, P., & Dolan, R. J. (2014). A computational and neural model of momentary subjective well-being. *Proceedings of the National Academy of Sciences*, *111*(33), 12252–12257. <https://doi.org/10.1073/pnas.1407535111>

Salkovskis, P. M., Hackmann, A., Wells, A., Gelder, M. G., & Clark, D. M. (2007). Belief disconfirmation versus habituation approaches to situational exposure in panic disorder with agoraphobia: a pilot study. *Behaviour Research and Therapy*, *45*(5), 877–885. <https://doi.org/10.1016/j.brat.2006.02.008>

Saris, I. M. J., Aghajani, M., Werff, S. J. A. van der, Wee, N. J. A. van der, & Penninx, B. W. J. H. (2017). Social functioning in patients with depressive and anxiety disorders. *Acta Psychiatrica Scandinavica*, *136*(4), 352–361. <https://doi.org/10.1111/acps.12774>

Saunders, J. B., Aasland, O. G., Babor, T. F., Fuente, J. R. de la, & Grant, M. (1993). Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption–II. *Addiction (Abingdon, England)*, *88*(6), 791–804. <https://doi.org/10.1111/j.1360-0443.1993.tb02093.x>

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of Internal Medicine*, *166*(10), 1092–1097. <https://doi.org/10.1001/archinte.166.10.1092>

Stringaris, A., Goodman, R., Ferdinando, S., Razdan, V., Muhrer, E., Leibenluft, E., & Brotman, M. A. (2012). The affective reactivity index: A concise irritability scale for clinical and research settings. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *53*(11), 1109–1117. <https://doi.org/10.1111/j.1469-7610.2012.02561.x>

Tomasello, M. (2014). The ultra-social animal. *European Journal of Social Psychology*, *44*(3), 187–194. <https://doi.org/10.1002/ejsp.2015>

Trivers, R. L. (1971). The evolution of reciprocal altruism. *The Quarterly Review of Biology*, *46*(1), 35–57. <https://doi.org/10.1086/406755>

Wickramaratne, P. J., Yangchen, T., Lepow, L., Patra, B. G., Glicksburg, B., Talati, A., Adekkanattu, P., Ryu, E., Biernacka, J. M., Charney, A., Mann, J. J., Pathak, J., Olfson, M., & Weissman, M. M. (2022). Social connectedness as a determinant of mental health: A scoping review. *PLOS ONE*, *17*(10), e0275004. <https://doi.org/10.1371/journal.pone.0275004>

Wittchen, H. U., Fuetsch, M., Sonntag, H., Müller, N., & Liebowitz, M. (2000). Disability and quality of life in pure and comorbid social phobia. Findings from a controlled study. *Eur Psychiatry*.